#4-99

Predicting the Equity Premium

June 1999
Revised August 1999

Amit Goyal
Anderson Graduate School of Management
University of California, Los Angeles

and

Ivo Welch
Anderson Graduate School of Management
University of California, Los Angeles

Finance Working Paper
sponsored by

CMS CAPITAL MANAGEMENT SCIENCES

Predicting The Equity Premium*

Amit Goyal† Ivo Welch‡

August 14, 1999

Abstract

Even though we confirm strong ability of the lagged dividend yield to predict equity premia *in-sample*, it has had no ability to predict *out-of-sample* in the 1946–1997 period. Dividend-yield regressions significantly outperformed the prevailing historical equity premium average as predictors only in two years (1973 and 1974). The reason for the discrepancy between in-sample and out-of-sample predictive ability is *parameter instability*, i.e., a time-varying correlation between the dividend yield and expected returns. A "changing market" model which suggests time-decay in the dividend yield coefficient dominates any other known models in forecasting the equity premium, either in-sample or out-of-sample, and predicts an equity premium of 3% to 5% as of 1999. But this model also predicts that the dividend yield has lost all marginal ability to forecast as of the late 1990's. Consequently, the view that the dividend yield predicts future equity premia as of 1999 must necessarily be exclusively based on a theoretical prior.

*This paper contains parts of an earlier, different paper entitiled "The Myth of Predictability," which focused on findings in Section 2 and the appendix. The paper's main data (variables and the predictions) is available from `http://linux.anderson.ucla.edu/research.html`. We thank Eugene Fama, Ken French, Will Goetzman, Charles Lee, Richard Roll, Pedro Santa-Clara, Eduardo Schwartz, Robert Stambaugh, and especially Luis Viceira and Tuomo O. Vuolteenaho for feedback. This paper is UCLA-Anderson working paper 4-99.

†Contact: `mailto:amit.goyal@agsm.ucla.edu`. US mail address: US mail address: Anderson Graduate School of Management at UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, Tel: (310) 825-2508.

‡Contact: `mailto:ivo.welch@anderson.ucla.edu`. US mail address: Anderson Graduate School of Management at UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, Tel: (310) 825-2508.

The dividend yield is commonly thought to predict stock returns. There is both a theoretical foundation and ample empirical evidence.

The most common theoretical illustration for why the dividend yield is likely to be related to future returns is based on a variant of the well-known dividend growth model. Under certainty,

$$P(t-1) = \frac{D(t)}{r-g}, \tag{1}$$

where P is the price, D the dividend, r the appropriate discount factor, g the growth rate of dividends ($r > g$), and t a time index. Inverting this equation,

$$r = g + \frac{D(t)}{P(t-1)} = g + \frac{(1+g) \times D(t-1)}{P(t-1)} \tag{2}$$

yields a relation between the expected rate of return and the dividend yield.[1]

Empirically, the use of the dividend yield to predict the equity premium (or stock returns) has a long tradition in finance (e.g., Dow (1920), Ball (1978)). More recently, Rozeff (1984), Shiller (1984), Campbell and Shiller (1988), Fama and French (1988) and Fama and French (1989) have reinvigorated this interest. (Cochrane (1997) provides an excellent survey of the equity premium literature.) Generally, the dividend yield is found to be statistically significant, especially over horizons longer than one quarter. This empirical regularity, that the dividend yield seems to predict equity returns, ranks among the more important findings of academic finance. For example, a citation search lists more than 150 published articles as of 1998 citing the Fama and French (1988) article. Among the follow-up literature are such high-profile papers as Lamont (1998), which adds earnings and scaled stock prices; and Lee, Myers, and Swaminathan (1999), which add more complex measures based on analyst and accounting valuation measures for the Dow-30. Some of these papers construct more complex measures which have data availability problems that restrict them to start their sample at about the post-war period. Their findings are

[1]The equity return prediction regressions are *not* designed to be formal tests of the dividend growth model. They only serve as backdrop to justify why dividend yields could be a relevant forecasting variable. Furthermore, an alternative to this efficient markets perspective is that the dividend yield predicts future stocks returns because prices swing predictably.

typically that the dividend yield is significant in at least some specifications, but less so in competition with their own variables. Some of these papers also find that their own variables outperform the dividend yield model in out-of-sample prediction. In turn, a number of theories have recently appeared that consider the impact of predictability. For example, Barberis (1998), Brennan, Schwartz, and Lagnado (1997), Campbell and Viceira (1998), Liu (1999), Lynch (1999), and Xia (1999) build models for asset-allocation based on equity premia processes that time-vary with the dividend yield.

After introducing the data in Section 1, our own paper begins in Section 2 by showing that, even though we can confirm good *in-sample* predictive ability for annual equity premia, the dividend yield has had poor *out-of-sample* forecasting ability. In out-of-sample tests, both the "dividend-yield conditional model" (regressions) and the "unconditional (mean) historical up-to-date equity premium" (moving average) models are estimated as rolling forecasts to predict one-year ahead equity premia. It is of course already "casually known" that statistical dividend-yield forecasting models performed poorly in the 1990's. It is less well known that this forecasting failure has a far longer history. Indeed, there are only two years (the 1973 and 1974 equity premia) in which dividend yields predict much better than the unconditional average. For dividend yields that use the ending period price (commonly denoted as $D(t-1)/P(t-1)$), the early 1950's also saw better performance by the conditional model. The unconditional model performed far better from the late 1950's through the early 1970's, and for $D(t-1)/P(t-2)$ post 1975 (this is delayed to the 1990's for $D(t-1)/P(t-1)$.) Over the entire sample period, taking the sum-total of all out-of-sample predictions from 1946–1997, the unconditional equity premium performed *better* than either dividend-yield model forecasts. Even over 1946–1990, there is no statistically significant ability of the dividend yield to predict better out-of-sample—and without 1973 and 1974, it cannot even predict *on average* better than the unconditional model. Thus, we conclude that the evidence that the equity premium varies with the dividend yield is tenuous: a market-timing trader could not have taken advantage of any dividend yield predictive ability. By assuming that the equity premium was "like it always has been," this trader would

have performed at least as well in most of the sample.[2]

In Section 3, we investigate why in-sample and out-of-sample tests provide such different results. We argue that the most likely cause is mis-specification: the dividend yield coefficient has been changing which renders out-of-sample forecasts unreliable. Even if the true relationship between dividends and future equity premia is positive, a plain linear dividend yield regression adds so much noise that it is not useful for prediction. We develop a statistical test for model stability based on out-of-sample performance, taking into account the serial autocorrelation in dividend-yields (pointed out in Stambaugh (1999) and Yan (1999)) and show that the best stable $D(t-1)/P(t-2)$ model that fits the in-sample evidence cannot deliver relative out-of-sample performance as poor as that observed in the 1946–1997 data. Even though the $D(t-1)/P(t-1)$ also has poor out-of-sample performance, it is not poor enough to reject the null hypothesis of model stability at conventional statistical significance levels.

Section 4 informally develops a reasonable alternative to the constant dividend-yield coefficient model. In this model, the market experienced a change in equilibrium return relations. We show that when the dividend-yield coefficient is allowed to decline deterministically (with the log of a simple year index), the in-sample *adjusted* R^2 predicting annual equity premia increases from 5.9% to 16.7% for $D(t-1)/P(t-2)$ based specifications and from 7.2% to 12.8% for $D(t-1)/P(t-1)$ based specifications. We believe that these two simple models offer higher in-sample adjusted R^2 when forecasting annual equity premia than any known competing predictive model of annual equity premia. Furthermore, although the $D(t-1)/P(t-1)$-based changing-market specification has inferior in-sample performance compared to the $D(t-1)/P(t-2)$-based changing-market model, it has superior out-of-sample performance. Indeed, the $D(t-1)/P(t-1)$ "changing market" model is the only model

[2]Appendix B reconciles our findings to those in Fama and French (1988), and shows that our conclusion is robust. Their out-of-sample period (1967–1986) just missed the poor dividend yield performance in the early 1960's and 1990's. Even in their period, the Fama-French conditional model did not outperform the unconditional benchmark out-of-sample in a statistically significant manner. Appendix C also shows that the poor out-of-sample predictability also applies to variables introduced by Lamont (1998) and Lee, Myers, and Swaminathan (1999).

we are aware of that succeeds in beating the unconditional model out-of-sample.

Finally, we argue that this change in return forecasting ability can be either because of a time-changing process in the persistence of the dividend yield, or due to investors adjusting their portfolios to try to take advantage of the dividend-yield forecasting ability (which in turn caused the dividend yield to lose its forecasting power), or both. We show that the persistence in dividend-yields has *increased* over the years. Although this increase in persistence can theoretically induce either an increase or a decrease in the dividend yield's ability to predict returns, given the estimated parameters, the increasing dividend-yield persistence is consistent with the observed decline in predictive power. In this interpretation, changes in the market dividend-yield process drove the decline in predictive ability. A related explanation for the dividend yield's decline is a "learning market" hypothesis, in which investors attempted to take advantage (via market-timing strategies) of the dividend yield's forecasting ability, and thereby drove its forecasting power to extinction. In this interpretation, investors' belief in an ability to use (published) academic research to earn higher returns during certain periods drove the decline in predictive ability.[3]

In sum, the best model describing the dividend-yield and equity premium data (adjusting for serial correlation in dividend-yields, which influences both the predictive and the auto-regression, as of 1999 is[4]

[3]This section also discusses the theoretical asset allocation models mentioned above, which assume a fixed linear relationship and learning by an individual who optimizes his asset-allocation as he learns the relationship over time. Such models incorrectly pursue a fixed target (coefficient), and thus end up recommending poor investment choices: they prescribe more aggressive investment in stocks as uncertainty about the dividend yield coefficient decreases, at least for U.S. data. Any strategies that increase the predictive importance of the dividend yield over time would of course have lead to even worse investment performance than fixed dividend-yield strategies. Still, if these models were to be extended to embed a moving target, they could be the first step in producing a dynamic general equilibrium model, in which the updating process itself alters equilibrium returns.

[4]The sample moments of this model estimated from 1926–1997 are

$$a(t) = -93.784 + 21.87 \times \log(t) \qquad c(t) = 3.778 - 0.887 \times \log(t)$$

$$b(t) = 19.043 - 3.868 \times \log(t) \qquad d(t) = 0.217 + 0.178 \times \log(t)\ .$$

$$EQP(t) = a(t) + b(t) \times \left[\frac{D(t-1)}{P(t-1)}\right] + e(t) \qquad \frac{D(t)}{P(t)} = c(t) + d(t) \times \left[\frac{D(t-1)}{P(t-1)}\right] + n(t)$$

$$a(t) = -85.39 + 20.75 \times \log(t) \qquad c(t) = 3.15 - 0.81 \times \log(t)$$

$$b(t) = +17.52 - 3.87 \times \log(t) \qquad d(t) = 0.313 + 0.170 \times \log(t)$$

Var[e(t)] = 357.1 Cov[e(t),n(t)] = -8.72 Var[n(t)] = 0.58

where $EQP(t)$ is the annual equity premium, D is the paid (reinvested) dividends, P is the price level on the value-weighted market portfolio, and t a year index which begins in 1926 as "0."

As of 1999, this model produces equity premia forecasts of about 5%,[5] not the theoretically unjustifiable (approx.) -5% to -8% predicted by traditional linear dividend-yield models. "Unfortunately," the changing-market model also estimates that the net dividend-yield influence is close to zero as of 1999 (unlike the standard approach which continues to indicate a statistically significant influence of dividend-yields).

The data evidence should suggest to an investor to realize (a) that simple linear dividend yield models have not helped predicting historically out-of-sample, and (b) that the best in-sample model indicates that dividend yields are not useful as of 1999. Yet, we still advise a reader to exercise caution in interpreting our results: First, we can only show that the data does not support a predictive ability for dividend yields; theoretical considerations (e.g., due to stronger priors) may allow an investor to continue believing in dividend-yields as a predictor for the future. Second, we view our evidence as a first step in modelling time-changing dividend yield coefficients—although our specification improves dramatically relative to the simple linear model, further research could relate changes in the dividend-yield coefficient to observable other variables, e.g. the business cycle. This issue is further explored in Section 5, in which we attempt to highlight some of the time-series aspects of the dividend-yield coefficient.

Fama and French (1988) ranks among the most influential papers of the last decade, so it is not surprising that a number of other papers have pointed out concerns in using the dividend yield to predict equity premia and stock returns

[5]The sample moments model produces a forecast of 3.67%, as reported in Table 5.

and/or introduced other variables. For example, Goetzmann and Jorion (1993) use a bootstrap to evaluate the in-sample predictive performance of coefficient estimates and find that the Fama and French (1988) coefficient estimates are upward biased. Nelson and Kim (1993) examine coefficient biases and come to similar conclusions. Goetzman and Jorion (1995) find that predictability in a longer sample (since 1872) is marginal and argues that these tests are influenced by survivorship bias. Hodrick (1992) finds that Hansen-Hodrick and Newey-West statistics are biased on longer than 1-year horizon. Stambaugh (1999) and Yan (1999) find that near-nonstationarity in the dividend yield biases the t-statistics and R^2. None of these entertains our simple out-of-sample naive benchmark comparison. Fama and French (1989) also use our naive benchmark, but their dividend forecast model even seems to outperform out-of-sample relative to their in-sample performance. (It is easy to miss this evidence, because the focus in Fama and French (1989) is the addition of fixed income variables to the dividend-yield.) Independently, Lee and Swaminathan (1998) find that the dividend yield has poor out-of-sample predictive ability in competition with their value-price ratio. After inclusion of their V/P measure, the dividend yield has no marginal explanatory power. Their more sophisticated model employing the V/P measure can beat a "static investment allocation" model, but only mildly so. Similarly, Lee, Myers, and Swaminathan (1999) find that, from 1963–1996, traditional market ratios had little (in-sample) predictive power. The closest paper to our own in pointing out poor out-of-sample power may be Bossaerts and Hillion (1999), which investigates more stringent model-selection criteria for data from a number of countries. Still, they find no out-of-sample predictability in a 6/90 to 5/95 hold-out sample, using $D(t-1)/P(t-2)$ as their forecaster. The closest paper to our own in pointing out the possibility of a changing market is Viceira (1997), who tests whether there is a structural break in the relation between the dividend yield and stock returns (but fails to detect one).

1 Data

Table 1 lists the data used in the paper. The data are well-known: value-weighted stock returns, with and without dividends, and the treasury bill yield. The data are available from 1926 through 1997. The derived series of interest are the equity premium, EQP, and the dividend yield, $D(t-1)/P(t-2)$ and $D(t-1)/P(t-1)$. The dividend yield is computed as the difference between the properly compounded value-weighted return with dividends and the properly compounded value-weighted return without dividends:

$$\mathsf{VWR}(t-1,t)-\mathsf{VWRX}(t-1,t) \equiv \frac{P(t)-P(t-1)+D(t-1,t)}{P(t-1)} - \frac{P(t)-P(t-1)}{P(t-1)} \equiv \frac{D(t-1,t)}{P(t-1)}. \tag{3}$$

Our definition de-facto reinvests dividends until the end of the year. We abbreviate flows (e.g., $D(t-1,t)$) quoting only the final period (e.g. $D(t)$). To compute $D(t)/P(t)$, we multiply by the market capitalization ration $P(t-1)/P(t)$ after adjusting for the fact that 1962 and 1972 saw the introduction of AMEX and NASDAQ.

[Insert Table 1: Data Sources]

[Insert Table 2: Descriptive Statistics]

Table 2 provides the descriptive statistics for the series. The properties of these series are well-known. The average log equity premium was 6.3% in our sample period, the average log dividend yield was 4.5%. Figure 1 plots the time series of our regressand (the equity premium) and our regressors (the dividend yield and changes therein). The latter makes it apparent that there is some nonstationarity in the dividend yield. The augmented Dickey-Fuller test (Dickey and Fuller (1979)) indicates that the dividend yield is likely to contain a unit-root (see Stambaugh (1999) and Yan (1999)). Thus, our paper also entertains changes in the dividend yield as a predictor, and introduces tests in Section 3 that explicitly adjusts for this nonstationarity.

[Insert Figure 1: Time Series Graphs]

2 In-Sample Fit and Out-of-Sample Prediction

2.1 In-Sample Fit

Table 3 correlates the equity premium with the lagged dividend yield and lagged dividend yield changes.[6] Our specifications differ slightly from earlier work (as earlier work does from one another), but our conclusions are mostly unaffected if, e.g., Fama and French (1988) specifications are used (see Appendix B). The sample dividend yield regressions confirm the findings in Fama and French (1988, Table 3). $D(t-1)/P(t-1)$ performs better as a predictor than the (perhaps more common) $D(t-1)/P(t-2)$. Table 3 also shows that, although the dividend yield is a nonstationary variable, changes in the dividend yield do not offer improved fit.

[Insert Table 3: Bivariate Regressions Predicting the Equity Premium (EQP) In-Sample]

2.2 Out-of-Sample Forecasts

[Insert Table 4: Properties of Forecast Errors Predicting The Equity Premium (EQP) Out-of-Sample]

Unfortunately, even a sophisticated trader could not have used the regression in Table 3 to predict the equity premium. A trader could only have used historical information to estimate a model and come up with a forecast. Consequently, Table 4 displays statistics on the prediction errors when the dividend yield model and an unconditional equity premium means forecast are estimated only with historical data. In each box, the first two data columns contain the in-sample prediction errors from the single full period regression model, as in Table 3. To do an out-of-sample comparison, we need an initial sample period to estimate coefficients. Thus, we chose (ex-ante) the post-war period (1946–97) as our out-of-sample window. The second two data columns display the in-sample 1946–1997 residuals' standard error from

[6] Differencing only the independent variable but not the dependent variable means that we are not testing the same model. Our purpose for entertaining this variable (for which we have no theory) will become clearer in the subperiod regression.

our single full-period 1926–97 regression. The final two data columns display the statistics of most interest: the performance of the out-of-sample rolling prediction errors for the 1946–1997 period. Again, each year, we use only available historical information to estimate the dividend yield regression. The regression coefficients are used to forecast the equity premium, and the statistics are over the sum-total of out-of-sample single year forecast errors. The out-of-sample benchmark and null hypothesis is that the next year's equity return is simply the same as the historical average, up to this date. This is denoted as **UNC** (unconditional [i.e., without dividend yield conditioning]).

Table 3 had indicated that annual equity premia are well-predicted in-sample. The top panel considers $D(t-1)/P(t-2)$. The first two data columns of Table 4 show that, when compared to the standard deviation of the overall equity premium series, the dividend yield regression model produces a meaningful 0.6% reduction (from 19.9% for the unconditional average to 19.3%) in the in-sample RMSE (root mean squared error), and an 0.4% reduction (from 15.3% to 14.9%) in the in-sample MAD (mean absolute deviation) in the 1926–1997 period. This holds equally true for the 1946–1997 period residuals, where the 1926–1997 estimated dividend-based model offers an 0.2% reduction in RMSE and an 0.3% reduction in MAD. Our shorter 1946–1997 period is not unusual in the ability of the dividend yield to shrink the in-sample variance in the context of our regressions. However, the final two columns of the table are surprising: When the dividend yield model is estimated only with historically available data and compared to the up-to-date historical mean, the dividend yield based model not only does not outperform the unconditional model, it actually *underperforms* it. The conditional dividend regression model has higher bias, a higher standard error, and thus a higher RMSE than the simple historical unconditional equity premium average. The bottom panel of Table 4 shows that although $D(t-1)/P(t-1)$ performs much better than its $D(t-1)/P(t-2)$ cousin, much (but not all) of its poor performance is due to the large bias of the dividend yield model: on average, the out-of-sample predictions of the dividend-yield models have simply been far more conservative than the historical mean. The dividend models also have a higher mean-absolute deviation (MAD) than the unconditional model.

In sum, for $D(t-1)/P(t-2)$, the difference of 1.3% per year on an RMSE basis and 1.1% per year on an MAD basis is sufficiently large to be economically meaningful to a market timer.[7] For $D(t-1)/P(t-1)$, the unconditional model predicts roughly as well as (but not worse) than the dividend-yield model. Our conclusion is that the in-sample performance is so different from the out-of-sample performance that one cannot implicitly combine the two measures to conclude that the dividend yield predicts the equity premium.

2.3 Statistical Significance (The Jackknife)

To assess whether the RMSE's are statistically significantly different, we run a simple *Jackknife*. To obtain a distribution of RMSE's *under the null hypothesis of equal performance of the two strategies*, we randomly reshuffle each of the 52 forecast errors from two models (the conditional and the unconditional model) into two new columns and compute the new RMSE difference to simulate a draw from the unconditional distribution. This procedure is repeated 10,000 times, and the *location* (i.e., the CDF) of the actually observed RMSE difference within this null distribution of RMSE differences is reported. The actually observed RMSE difference of 1.3% for $D(t-1)/P(t-2)$ lies roughly at the 83rd percentile under the null, indicating only weak statistical significance (i.e., at the 20% level on a one-sided test). It is assuring that the dividend yield model does not predict statistically significantly *worse* than the null hypothesis at conventional significance levels. There is no meaningful difference between the out-of-sample forecasts of $D(t-1)/P(t-1)$ and the unconditional model.

We can safely conclude that an investor could as well have used the historical average equity premium instead of dividend-yield forecasts, and performed at least as well. Market participants without a strong prior should not have believed that

[7] Bossaerts and Hillion (1999) focus on $D(t-1)/P(t-2)$ and find failure to predict out-of-sample in the last 5 years. Our own paper shows that failure of $D(t-1)/P(t-2)$ to predict out-of-sample is more systematic, going back to (at least) 1946 (see also Appendix B). Further, if Bossaerts and Hillion (1999) had entertained $D(t-1)/P(t-1)$, the out-of-sample performance of the dividend yield model would have been better.

they could use the dividend yield in a productive manner to time the market. The dividend yield could not reliably predict the equity premium out-of-sample, even when in-sample regressions indicate it could. The burden of proof must lie with the claim that the dividend yield can predict the equity premium.

One might object to our findings based on issues of statistical power. However, it is unclear what modifications one should make to increase power. Both the null hypothesis (the unconditional mean) and the alternative hypothesis (the regression model) are clearly defined in the literature, as are the metrics on which they are compared (Bias, RMSE, MAD). Alternative, more powerful models are not at debate here. (We introduce a better model in Section 4.) It is the simple linear regression model which has been proposed in the academic finance literature as an improvement over the null hypothesis, and its evidence relied on the very same data sample used in our paper. Our evidence does not dispute that the dividend yield might in truth predict the equity premium (e.g., because we have too few observations or because our prior is too weak) or that we use the wrong simple linear specification. Instead, our paper claims that, as of today, taken as a whole, the data by itself suggests that the profession should not presume that the dividend yield model can predict the equity premium in the simple linear fashion usually presumed.

2.4 Alternative Specifications

We also tried numerous variations. First, we tried predicting on different horizons (monthly, quarterly, multi-yearly), although annual horizons seem to have been generally agreed to have the least statistical problems and the best or close-to-best performance. Sometimes, other frequencies improve the relative performance of the unconditional model, sometimes they improve the relative performance of the dividend-yield model. Under no frequency did we find the dividend yield model to outperform in predicting at a halfway statistically significant manner. Second, we used simple instead of log returns and yields. For $D(t-1)/P(t-2)$, on an out-of-sample basis, the conditional prediction had an RMSE of 18.4%, the unconditional prediction had an RMSE of 17.4%. Again, the unconditional model beats

the dividend yield model. Third, we tried standardized forecasts to see if the regressions/means could identify years ex-ante in which it was likely to perform unreliably. (In other words, we used the regression prediction standard error to normalize forecast errors.) Again, the unconditional model (its forecast also standardized by its standard deviation) beat both versions of the conditional model. Fourth, we tried a convex combination of the dividend yield model prediction and the unconditional prediction. Such a "shrunk dividend yield" model does not produce meaningfully better forecasts than the unconditional model alone. Fifth, we tried forecasting with the Stambaugh (1999) correction for high serial correlation in the dividend yield. This worsens the out-of-sample performance, even though the average dividend-yield coefficient decreases in the $D(t-1)/P(t-2)$ specification (by 0.70 on average) and increases in the $D(t-1)/P(t-1)$ specification (by 1.34 on average). For $D(t-1)/P(t-2)$, the RMSE increases from 17.3 to 17.6; for $D(t-1)/P(t-2)$, the RMSE increases from 16.2 to 18.6. *Both specifications reach statistical significance (using the Jackknife) at the 10% one-sided level in the dividend-yield's underperformance relative to the unconditional historical average.* Sixth, we tried various changes in our definitions in line with Fama and French (1988). Most are minor and explained in Appendix B. Figure 2 displays the single major difference: the choice of out-of-sample window. A positive slope indicates years in which the dividend-yield model outperformed the unconditional historical historical equity premium average, a negative slope indicates years in which the unconditional model outperformed. The figure shows that the predictive regressions were better from 1946 to 1957 and in 1973 and 1974. From 1956 to 1965 and the 1990s have been particularly poor years for out-of-sample prediction, although $D(t-1)/P(t-2)$ also performed poorly ever since 1975. The Fama-French out-of-sample period began just after the dividend-yield model had ended a 10-year poor run, and ended just three years before $D(t-1)/P(t-1)$ began deteriorating.[8] The figure also shows

[8]Fama and French (1988) and Fama and French (1989) use estimation periods of 30-years to obtain an out-of-sample estimation period from 1967 to 1986 and 1967 to 1987, respectively, which avoids some high-variance returns in the 1930's. As FF point out, an investor may have recognized that the post-war period was different enough from the pre-war period to avoid using an estimated dividend yield regression to predict equity premia prior to 1967. Similarly, the 1990's poor out-of-sample performance occurred after the Fama and French (1988) paper was written—and we know that the in-sample relationship has recently declined.

that it is only the years of 1973 and 1974 that "rescue" the dividend-yield model (at least the $D(t-1)/P(t-2)$ model).[9]

[Insert Figure 2: Cumulative Out-of-Sample Relative Out-of-Sample Sum-Squared Error Performance]

Appendix C replicates out-of-sample forecasts for variables proposed by Lamont (1998) and Lee, Myers, and Swaminathan (1999). In sum, these models have been similarly unable to beat the simple historical equity premium average in an out-of-sample forecasting horserace in a statistically or economically significant fashion.

In sum, variations on the specification and variable did not produce instances which would lead one to believe that dividend yields or other variables can predict equity premia in a meaningful way. The conditional dividend yield $D(t-1)/P(t-2)$ regression model predicts no better than the historical unconditional equity premium at least since 1946. The conditional dividend yield $D(t-1)/P(t-1)$ regression model predicts no better than the historical unconditional equity premium in an overall sense at least since the early 1990's. As of 1999, the data no longer support the view that dividend yields were an effective forecasting tool over the 1946–1997 period. It is not likely that there is a simple model variation which can induce superior out-of-sample performance for the dividend-yield model.

Finally, it is also important to recognize that different published papers may have come to slightly different results, depending on how they lag the price deflator. For example, Bossaerts and Hillion (1999) employ the more common $D(t-1)/P(t-2)$. Consequently, they find much worse out-of-sample performance in their 5-year out-of-sample period than if they had used $D(t-1)/P(t-1)$. Fama and French (1988) report both measures, but emphasize the better performance of the $D(t-1)/P(t-1)$ measure.

[Insert Figure 3: Different Lengths of Estimation Windows To Predict the Equity Premium (EQP) Out-Of-Sample]

[9]This conclusion also emerges when one begins estimation with 1945 data. In particular, the dividend-yield models are "rescued" from outright *inferior* performance only by the years of 1973 and 1974. To view the equivalent figure, please refer to *http://linux.agsm.ucla.edu/academics/ooscum45.gif.*

A number of papers (e.g., Barberis (1998)) work out how an individual investor should have updated their investment decision while learning about the dividend-yield coefficient. So we wondered how long a memory a market participant needed to make a reasonable prediction of the future equity premium using only the past average. Figure 3 graphs the RMSE when the estimation period used to forecast out-of-sample is not "from inception," but limited to a given number of years. As before, we begin estimating out-of-sample forecasts in 1946. When not enough years are available (because the beginning of our window wants to start pre-1926), the estimation period begins in 1926. By setting the window to the length of the longest possible period (72 years), i.e. by looking at the right edge of each figure, we obtain the RMSE reported in Table 4.

The figure shows that a memory of the most recent 5 years did a good job for the unconditional model—enough to beat the dividend regression models. A memory of 10 years did a better job in predicting the equity premium. To be halfway reliable, the regression estimation period requires at least 10 years of data. At no regression estimation window size does the conditional $D(t-1)/P(t-2)$ model reliably beat the unconditional model; but for certain estimation windows (27 to 36 years), $D(t-1)/P(t-1)$ can beat the unconditional historical average by a small margin.

In sum, the relative performance of either the conditional or the unconditional prediction is not driven by an unusual property of our ever-extending these from inception.

2.5 Investor Utility Losses

The changing dividend-yield coefficient added sufficient noise to the dividend yield to forestall any attempts at using it for predictive purposes. How much would an investor have lost who could have used the historical average as the best predictor of the future equity premium, but who instead was forced to optimize with respect to the conditional dividend yield model? We answer this question in a naive setup. We assume a representative investor with exponential utility, a risk-free asset and

the market only. It is straightforward to show[10] that the ratio

$$\frac{EU_{\text{unconditional}}}{EU_{\text{conditional div}}} = e^{\left[\frac{(\mu_1+\mu_2-2\mu_m)\times(\mu_1-\mu_2)}{2\sigma_m^2}\right]} \tag{4}$$

where $EU_{\text{unconditional}}$ is the expected utility of an optimizing investor living in a world in which the historical mean is the best predictor of the equity premium and $EU_{\text{conditional div}}$ is the expected utility of this aware investor when forced into the same strategy that an optimizing investor follows who believes in a dividend-yield timing strategy. Note that the ratio does not depend on either initial wealth and risk aversion and cannot be greater than 100%. Our interest is in magnitudes here.

If forced into following decisions from a model using the simple linear $D(t-1)/P(t-2)$, the forced investor would have received only 91.7% of his (negative) expected utility on average (standard deviation: 0.10%; range from 0.45% to 100%). To compute the ratio of certainty equivalents, one needs to assume an absolute risk-aversion coefficient. If it is one, the "unconditional investor" would have given up about 9% of his wealth in order not to be forced into the investment strategy of the "conditional investor." This investor would have sacrificed a significant amount of wealth not to be forced to follow the strategy that the conditional average prescribed. Clearly, inference differences about expected stock returns can represent an economically significant difference in *ex-ante* welfare. If forced into following decisions from a model using $D(t-1)/P(t-1)$, there is obviously very little difference between the unconditional and the conditional model.[11]

2.6 Conclusion

Given our findings, what should an investor do? One might want to argue that the standard linear dividend yield regression still have meaning in a Bayesian sense and that investors should use historical regressions, even if they failed to predict

[10]Details available upon request.

[11]Breen, Glosten, and Jagannathan (1989) focus is on the economic significance in the context of a model that times the market out-of-sample, using the short-term interest rate as a predictor.

out-of-sample. Why put 100% weight on the historical average equity premium, just because it predicted better out-of-sample than the conditional model in the past? After all, a Bayesian updater would still use all available data up-to-date, run a regression, and use it. An analogy is a situation in which a medical study finds a surgery procedure to be effective in 1,000 patients, and the fact that it fails in the next 100 (out-of-sample) patients does not mean that the procedure should be abandoned. Instead, patient 1,101 should benefit from the experience of all 1,100 patients up-to-date. Unfortunately, this analogy is flawed. Failure in the 100 out-of-sample patients is an unlikely scenario, and a statistician should learn from consistent out-of-sample failure that the model is misspecified. A better analogy than surgery might be penicillin: when historical regressions consistently mispredict treatment effectiveness out-of-sample, we should conclude that the most likely cause is that bacteria are becoming resistant. Of course, resistance implies different forecast and treatment implications. The remainder of this paper focuses on the "changing" of the market.

3 A Simple Test For Parameter Instability

Given the poor out-of-sample performance, our first question is how an investor should view the out-of-sample misprediction evidence in evaluating the linear dividend-yield models. We thus develop a simple test for model stability in the dividend-yield prediction context. We must adjust for the fact that when the dividend yield is almost a random walk, it can bias the estimated dividend yield coefficient, as pointed out by Stambaugh (1999) and Yan (1999). For our own purposes of testing for model instability, we substitute computer speed for the complexity of *small-sample* out-of-sample econometrics. Our goal is to estimate an underlying model of the form

$$\begin{aligned} \mathrm{EQP}(t) &= x_1 + x_2 \times \mathrm{DVY}(t-1) + \epsilon_E(t) && (5) \\ \mathrm{DVY}(t) &= x_3 + x_4 \times \mathrm{DVY}(t-1) + \epsilon_D(t) && (6) \\ \begin{pmatrix} \epsilon_E(t) \\ \epsilon_D(t) \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x_5 & x_6 \\ x_6 & x_7 \end{pmatrix} \times 10^{-3}\right] && (7) \end{aligned}$$

where EQP is the equity premium and DVY is the dividend yield, either $D(t-1)/P(t-1)$ or $D(t-1)/P(t-2)$. (As before, EQP and DVY are quoted in logs.) We want to match (using the same functional specification) the empirically observed 1927–1997 data sample moments:

Dividend Yield	x_1	x_2	x_3	x_4	x_5	x_6	x_7
D(t-1)/P(t-2)	-0.115	3.949	0.014	0.685	37.233	0.850	0.105
D(t-1)/P(t-1)	-0.144	4.977	0.007	0.830	36.714	-0.910	0.051

One can of course not use the data moments as best model estimates. It is well-known that if the true x_4 is close to 1, the sample x_4 is biased downward; and Stambaugh (1999) and Yan (1999) suggest that inference on x_2 is similarly biased if the true (not sample!) x_4 is close to 1. Our own goal is not to obtain inferences (i.e., significance levels) about x_2, but to test if the best stable model that fits the 1927-1997 data can generate poor out-of-sample performance in line with that observed in the real empirical data. With seven unknowns and seven moment conditions to

match, the underlying economy which, when simulated 100,000 times, produces average[12] moments matching those observed in the real 1927–1997 data is

$$\begin{aligned}
\text{EQP(t)} &= -0.135 + 4.298 \times \text{D(t-1)/P(t-2)} + \epsilon_E(t) && (8)\\
\text{D(t)/P(t-1)} &= +0.012 + 0.734 \times \text{D(t-1)/P(t-2)} + \epsilon_D(t) && (9)\\
\begin{pmatrix} \epsilon_E(t) \\ \epsilon_D(t) \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 39.344 & 0.891 \\ 0.891 & 0.116 \end{pmatrix} \times 10^{-3}\right] && (10)
\end{aligned}$$

for $D(t-1)/P(t-2)$ and

$$\begin{aligned}
\text{EQP(t)} &= -0.107 + 3.946 \times \text{D(t-1)/P(t-1)} + \epsilon_E(t) && (11)\\
\text{D(t)/P(t)} &= +0.004 + 0.888 \times \text{D(t-1)/P(t-1)} + \epsilon_D(t) && (12)\\
\begin{pmatrix} \epsilon_E(t) \\ \epsilon_D(t) \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 37.815 & -0.942 \\ -0.942 & 0.052 \end{pmatrix} \times 10^{-3}\right] && (13)
\end{aligned}$$

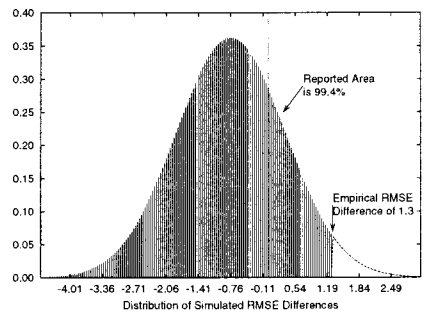
for $D(t-1)/P(t-1)$. For brevity, call each of these estimated systems a "true economy." The systems clearly display the dividend-yield predictive coefficient biases discussed in Stambaugh (1999) and Yan (1999), and a rough over-the-envelope calculation shows that its magnitude is very close to their predictions.[13]

For the $D(t-1)/P(t-2)$ dividend yield model: Using the "true economy," we simulate a million draws of 72-year data realizations. For each of these million draws of 72-year histories, we can compute 52 out-of-sample predictions for both a conditional and an unconditional model to compute a relative RMSE, and determine how frequently the conditional $D(t-1)/P(t-2)$ model underperforms by the 1.3% (17.3%–16.0%) observed in the real 1946–1997 data. We find that the simulated RMSE-difference distribution has a mean of -0.76% (in favor of the condi-

[12]Because the parameters are normally distributed in the limit (as more and more computer realizations are drawn to match), this can be interpreted as equivalent to the maximum likelihood estimate.

[13]It is comforting that this "true economy" also produces unconditional means and variances for dividend-yield and equity premium 72-year histories that are very close to those observed in the 1927–1997 data.

tional model) predicting out-of-sample, a median of 0.55%, and a standard deviation of 1.1%. The observed +1.3% realization lies at the 99.4th percentile, meaning that it is highly unlikely that a stable dividend-yield model with the observed in-sample moments could have generated such poor out-of-sample performance relative to the unconditional model. We also arbitrarily dropped the true x_2 coefficient from 4.298 to 2.149, holding all other true economy parameters constant. Because the dividend yield now has less predictive power, it would be less surprising to see the unconditional model outperform in such a world.



Indeed, on average in the simulations, the conditional model now showed about the same out-of-sample performance as the unconditional model (mean RMSE-difference: -0.07%; standard deviation 0.6%). However, the observed +1.3% RMSE underperformance in the 1946-1997 data still lies only at the 98.7th percentile in the simulated distribution. Even a model with half the predictability of the dividend yield cannot generate such poor out-of-sample performance reasonably often.

For the $D(t-1)/P(t-2)$ dividend yield model: The evidence does not allow rejecting the null hypothesis of stability at conventional significance level. The empirically observed 0.22% underperformance of the dividend-yield model relative to the unconditional historical average model lies at the 83rd percentile under the null hypothesis. (The null distribution's mean/median of RMSE is -0.5%, with a standard deviation of 0.86%.)

In sum, we believe that the evidence rejects the hypothesis that the observed 1927-1997 data is drawn from a system of the form that is described by equations (8-10), where the predictor is $D(t-1)/P(t-2)$. Such a system cannot generate relative out-of-sample underperformance as poor as that observed in the data. However, although there are hints that the $D(t-1)/P(t-1)$ model (eqs. 11-13) is unstable, we cannot reject the null hypothesis of stability at conventional significance levels.

4 A "Changing Market" Specification

4.1 Implications

Given that the basic model may not be stable, are there better models to predict the equity premium? There are at least two alternatives, both of which can explain changes in the predictive ability of dividend yields:

1. One alternative theory (with a flavor of market inefficiency) may explain how the market became "resistant" to dividend yield forecasts. As investors learned about the timing power of dividend yields, their collective efforts to exploit it reduced the power of the dividend yield. This is the familiar process that is commonly used to justify "rational expectations," albeit at a non-instantaneous speed.[14] Furthermore, if an individual investor performed Bayesian updating without allowing the dividend yield coefficient to change (decline) over time (e.g., as in Barberis (1998)), she would impose an incorrect constraint (a constant coefficient), and as a result would consistently make poor decisions. Ironically, if the "learning market" hypothesis is true, Fama and French (1988) may have contributed to the demise of its main finding, even though the dividend-yield coefficient seems to have declined for a long time before the 1990's.
2. Changes in the persistence of dividend yields induced changes in the relation between dividend yields and expected returns. This will be further examined in Subsection 4.2.

Both hypotheses suggest "changing market" conditions, which can be investigated by estimating

$$EQP(t) = a(t) + b(t) \times [DVY(t-1)] + e(t) \tag{14}$$

[14]This could be due to uncertainty in the prior; even though a full-process-knowledge investor would have considered the market inefficient, limited-process-knowledge investors may not have. Of course, we do not have a formal theory, so we cannot answer such interesting questions of how the dividend yield came to have predictive ability in the first place, what the functional relation should be, or how fast learning should take place. Even without such a formal theory, our evidence of much better predictability remains interesting in itself.

in the context of three models:

Coefficient	Model M1	Model M2	Model M3
$a(t)$	a	$c + d \times t$	$c + d \times \log(t)$
$b(t)$	b	$e + f \times t$	$e + f \times \log(t)$

where $t \in [1...72]$ is a simple year index. Model M1 is the ordinarily estimated equation. Model M2 permits the coefficient to change linearly each year. Model M3 permits the coefficient to change less and less each year. The standard dividend-yield model is nested in the testable restriction and $d = f = 0$. Substituting back, equity premia are now a function of lagged dividends and lagged dividend-year cross-terms. Clearly, these are not sophisticated models, and no claim is made that they provide the best process for time-varying coefficients. This specification is used because it captures the spirit of a "changing market" and because lends itself immediately to a simple OLS regression. Market-changing causes an omitted variables problem in the standard framework. In our case, it would have manifested itself in out-of-sample performance that is poorer than a model that restricts d and f to be zero, which illustrates why a Bayesian investor who did not allow the in-sample regression coefficient to change would come to incorrect conclusions.

On a more pragmatic level, a reader who does not believe in a "changing market/Bayesian updating" hypothesis but who is willing to concede that we did not run data snooping regressions (we did not!) should still be interested in whether such exceedingly simple specifications above can perform statistically significantly better than the simple linear specification.

[Insert Table 5: Time-Varying Dividend-Yield Coefficients]

Table 5 shows that including the year cross-terms roughly *doubles or triples* the explanatory power of the full in-sample regressions, depending on whether the dividend yield is presumed to have changed by a fixed or by a declining amount each year. Such a dramatic increase in forecast power is itself interesting to academics searching for the best in-sample predictive description of expected returns.

For the $D(t-1)/P(t-2)$ specification, the adjusted r-square reaches 16.7%, up from 6% for the simple linear specification. *We believe this "changing-market*

specification" to be the "best" in-sample forecaster of annual expected returns for the 1926-1997 period known today, where "best" means highest adjusted r-square. For the $D(t-1)/P(t-1)$ specification, the addition of the time-index cross-terms increases the adjusted r-square from 7.2% to 12.8%. Yet, although the in-sample statistics give the nod to $D(t-1)/P(t-2)$ over $D(t-1)/P(t-1)$, the out-of-sample statistics favor the latter model. Indeed, while the $D(t-1)/P(t-2)$ changing market specification improves the out-of-sample performance only slightly (from 17.31% to 17.06%) and not enough to beat the unconditional RMSE of 15.97%, the $D(t-1)/P(t-1)$ specification produces an out-of-sample RMSE of 15.55%, which *beats* the RMSE of the unconditional model. As far as we are aware, it is the only model that succeeds on this dimension as of 1999.

The "changing market" $D(t-1)/P(t-1)$ dividend yield model has some further desirable characteristics. First, not surprisingly, when we repeat our out-of-sample model stability test allowing both the equity premium forecast and the dividend yield regression to have time-varying log coefficients, one cannot reject the null hypothesis of stability for either the $D(t-1)/P(t-2)$ or $D(t-1)/P(t-1)$ model.[15] Second, the last two columns of Table 5 show that an investor's forecast of 1999's equity premium given a dividend yield of about 1.5% changes the forecast equity premium from an unpalatable -7% to -9% in the M1 specification to 12.3% in the M3 specification for $D(t-1)/P(t-2)$ and to 3.7% in the M4 specification.[16]

It is easy to compute the implied current marginal influence of the dividend-yield in predicting equity premia. The (biased) sample moments suggests an influence of $22.9 - 5.587 \times 4.3 \sim -1.1$ for $D(t-1)/P(t-2)$ and $19.0 - 3.9 \times 4.3 \sim 2.4$ for $D(t-1)/P(t-1)$. When corrected for biases (allowing the dividend yield auto-process to change over time, too), the equivalent implied coefficients change to

[15]The best model estimates, given the sample data, is $EQP(t) = -1.207 + 23.792 \times D(t-1)/P(t-2) + 0.300 \times \log(t) - 5.619 \times \log(t) \times D(t-1)/P(t-2)$, and $D(t)/P(t-1) = 0.040 + 0.188 \times D(t-1)/P(t-2) - 0.007 \times \log(t) + 0.160 \times D(t-1)/P(t-2) \times \log(t)$; and $EQP(t) = -0.854 + 17.520 \times D(t-1)/P(t-1) + 0.208 \times \log(t) - 3.867 \times \log(t) \times D(t-1)/P(t-1)$, and $D(t)/P(t) = 0.032 + 0.313 \times D(t-1)/P(t-1) - 0.008 \times \log(t) + 0.170 \times D(t-1)/P(t-1) \times \log(t)$. (Variance-covariance matrices are not reported). See also the next subsection.

[16]The 12.3% in the $D(t-1)/P(t-2)$ out-of-sample forecast is larger than the historical equity premium unconditional average, because model M3 also considers how the historical equity premium has increased over time.

-0.30 and $+0.93$, respectively. This is much smaller than the fixed linear dividend-yield coefficients of the M1 model (of about $+4/+5$); and, with such a low magnitude, none of the 1999 dividend yield coefficients is statistically significant from zero in the changing-market model.

[Insert Table 6: Time-Varying Dividend-Yield Coefficients as of 1990]

There is one question of historical interest: To what extent are our results surprising from the perspective of an observer standing at the beginning of the 1990s (i.e., right after the publication of the Fama and French (1989) paper). Table 6 produces estimates of models M1 and M3 when the sample ends in 1990. There is no meaningful difference for the $D(t-1)/P(t-2)$ results: all vital statistics are practically the same as those when the sample ends in 1997. (Not reported, our moment tests can reject M1 model stability at the 3rd percentile in 1990.) Consequently, an investor standing in 1990 could have concluded that market conditions were changing, that there was a better in-sample model, and that out-of-sample performance was rather problematic. Clearly, our $D(t-1)/P(t-2)$ results are not driven by the experience of the last few years.

For the $D(t-1)/P(t-1)$, however, the inference of a 1990 investor would have been more difficult: Although the in-sample adjusted r-square of the "changing market" model was slightly higher than that of the plain linear dividend-yield model, its out-of-sample performance was actually slightly worse. However, even on out-of-sample performance, there was more statistical probability that the changing model beat the unconditional historical average despite its worse RMSE. (This is caused by a tighter distribution of RMSE differences under the NULL hypothesis.) For $D(t-1)/P(t-1)$, a fair interpretation of the data is that an investor in 1990 could have suspected that dividend-yield coefficients are subject to log-decline, and that this suspicion was strengthened in the 1990's.

4.2 Are Dividend Persistence Changes Responsible?

One cause of the "changing market" may be changes in the persistence of the dividend yield process. The in-sample OLS estimates in a "changing persistence" regression specification are

$$\frac{D(t)}{P(t-1)} = 0.045+0.091\times\left[\frac{D(t-1)}{P(t-2)}\right]-0.009\times\log(t)+0.173\times\left\{\left[\frac{D(t-1)}{P(t-2)}\right]\times\log(t)\right\}+e(t) \quad (15)$$

and

$$\frac{D(t)}{P(t)} = 0.038+0.217\times\left[\frac{D(t-1)}{P(t-1)}\right]-0.009\times\log(t)+0.178\times\left\{\left[\frac{D(t-1)}{P(t-1)}\right]\times\log(t)\right\}+e(t) \quad (16)$$

where $t \in [1..72]$. Using Newey-West t-statistics, the 0.09 coefficient on $D(t-1)/P(t-1)$ is insignificant, indicating that early in our sample period, $D(t)/P(t-1)$ was almost iid. By 1999, however, the dividend yield process had developed its near unit-root characteristics ($0.09+0.173\times4.28 \sim 0.83$ and $0.217+0.178\times4.28 \sim 0.98$). Thus, OLS estimates are biased in small-sample time-series regressions. Repeating our moment-matching procedure, the best underlying model to fit the sample moment estimates are

$$\frac{D(t)}{P(t-1)} = 0.040+0.188\times\left[\frac{D(t-1)}{P(t-2)}\right]-0.007\times\log(t)+0.160\times\left\{\left[\frac{D(t-1)}{P(t-2)}\right]\times\log(t)\right\}+e(t) \quad (17)$$

and

$$\frac{D(t)}{P(t)} = 0.032+0.313\times\left[\frac{D(t-1)}{P(t-1)}\right]-0.008\times\log(t)+0.170\times\left\{\left[\frac{D(t-1)}{P(t-1)}\right]\times\log(t)\right\}+e(t) \quad (18)$$

Note that the dividend-yield term is downward biased, the log-time term is unaffected, and the log-time-dividend-yield cross-term (the term of interest) is slightly upward biased.[17]

Is this increase in dividend yield persistence consistent with the decline in its ability to predict future equity premia? Appendix A shows that the theoretical effect can be ambiguous: when dividend-yield persistence is low, increases in persistence can induce increases in predictive ability; however, when dividend-yield persistence is already high, further increases in persistence can induce decreases in predictive

[17]These underlying models suggests that while $D(t)/P(t-1)$ is highly autocorrelated, $D(t)/P(t)$ is likely to be a random walk as of 1999 ($0.313 + 0.17 \times 4.28 \sim 1.0$).

ability. Indeed, the estimated dividend-yield persistence level was such that it could theoretically have induced an increase in the dividend-yield's predictive ability in the first 5-10 years of our sample only, and a decrease in the dividend-yield's predictive ability in the remaining 60 years. When we split the sample into these two periods, we indeed observe a time-increase in the dividend-yield's predictive ability in the first 5-10 years and a time-decrease in the remaining 60 years. However, there are two few observations to draw final conclusions. For the dividend-price ratio $[D_t/P_t]$ increases in the dividend-price ratio should unambiguously cause decreases in the ability of dividend-price ratios to forecast equity premia, again consistent with our findings.

Further, Appendix A suggests that one can decompose changes in the predictive coefficient itself into changes in the persistence of the dividend-yield and into changes in the ability of the dividend yield to predict future dividends. This suggests the following procedure: Using δ as abbreviation for the dividend-price ratio, $D(t)/P(t)$, and d as abbreviation for $D(t)$, we first estimate the rolling regressions

$$r_u = \alpha_{1t} + \beta_{1t}\delta_{u-1} + e_{1u} \quad \forall u = 1, 2, ..., t \tag{19}$$

$$\delta_u = \alpha_{2t} + \beta_{2t}\delta_{u-1} + e_{2u} \quad \forall u = 1, 2, ..., t \tag{20}$$

$$\Delta d_u = \alpha_{3t} + \beta_{3t}\delta_{u-1} + e_{3u} \quad \forall u = 1, 2, ..., t \tag{21}$$

Appendix equation 24 suggests that one can decompose

$$\Delta\beta_{1t+1} = -\kappa\Delta\beta_{2t+1} + \Delta\beta_{3t+1} \tag{22}$$

where κ can be calibrated to be about 4% for U.S. data. (See the Appendix.) That is, parameter variation in the predictive coefficient can be due to parameter variation in the dividend yield process or in the dividend yield vs. dividend growth relation. Using these equations, we can run a variance decomposition for US data.[18] Of the

[18]One can also decompose the out-of-sample error into in-sample error and parameter instability. Using the equations in the text, one obtains a forecast $r^{f1}_{t+1} = \alpha_{1t} + \beta_{1t}\delta_t$. The coefficients $\{\alpha_{it}, \beta_{it}\}, i = 1, 2, 3$ are updated when information at time t+1 becomes available. This gives us another forecast based on updated coefficients as $r^{f2}_{t+1} = \alpha_{1t+1} + \beta_{1t+1}\delta_t$. Using these forecasts,

variation in recursive $\Delta\beta_{1t+1}$, $\Delta\beta_{2t+1}$ can explain 49.9% in a univariate specification, while $\Delta\beta_{3t+1}$ can explain about 6.5%.

Although one may be tempted to interpret persistence changes in dividend yield to be indicative of "statistical process changes" and changes in the dividend-yield dividend-growth relation to be indicative of "market learning," this is probably incorrect. The dividend-yield itself is an endogenous variable, and market learning could be responsible for process changes thereof. Unfortunately, we have little more to say here. Most of the relevant empirical literature treats the dividend-yield as an exogenous variable (as do we). We were surprised that we could not find a theory for the underlying processes for discount rates and dividends which are consistent with the observed random-walk in dividend-yields. Aside from one rather unappealing specification in which discount rates themselves follow a random walk, it is not even clear to us whether a random walk in dividend yield or dividend-price ratio is consistent with *any* efficient markets explanation.

4.3 Investor Learning Versus Market Learning

Before closing, we must emphasize that there is a fundamental difference between updating on an assumed stable linear D/P coefficient (e.g., as in the asset allocation models cited in the introduction; Barberis (1998) in particular entertains out-of-sample individual investor (not market!) learning over time), and systematic changes in the model coefficient (i.e., model misspecification) itself. Both issues allow for poorer out-of-sample performance than in-sample performance. But in the former case, one would expect not to see exceedingly poor out-of-sample fore-

---

we define two errors as $eos_{t+1} = r_{t+1} - r^{f1}_{t+1}$ and $eis_{t+1} = r_{t+1} - r^{f2}_{t+1}$. In other words,

$$\begin{aligned} r_{t+1} &= \alpha_{1t} + \beta_{1t}\delta_t + eos_{t+1} \\ r_{t+1} &= \alpha_{1t+1} + \beta_{1t+1}\delta_t + eis_{t+1} \end{aligned}$$

Differencing gives us

$$eos_{t+1} = \Delta\alpha_{t+1} + \Delta\beta_{1t+1}\delta_t + eis_{t+1}$$

By themselves, $\Delta\alpha_{t+1}$ can explain 4.5% of the variation in eos_{t+1}; $\Delta\beta_{1t+1}\delta_t$ can explain 12.9%; and the in-sample error can explain 99.8%.

casts and expect to see the forecast and investment allocation decisions to improve year by year, and—unless the estimated coefficient on D/P was exactly zero—an investor should still use the D/P model estimates for his investment choice today.[19] The poor out-of-sample performance would have to be dismissed as an (albeit statistically) unlikely fluke. In the latter case, statisticians would not be surprised by poor and worsening out-of-sample performance and would dismiss the estimated linear dividend yield regression model coefficients as inadequate for purposes of (out-of-sample) investment allocation purposes today.

Still, the aforementioned asset-allocation model can become very useful, if they can embed a time-varying relationship. This would allow one to solve for a rational expectations like equilibrium, in which investors make asset-allocation decisions in a general equilibrium framework, in which their decisions change the dynamics of the underlying stock return model itself.

4.4 Perspective

Although the above evidence is interesting in itself—producing better in-sample and out-of-sample models than hitherto in the literature—there are many reasons to be cautious. First, our finding could be due to specification search.[20] We tried one additional restriction $b(t) = \gamma_0 + \gamma_1 \times b(t-1)$ (where $b(t)$ denotes the dividend-yield predictive coefficient at time t) which did not fit as well, so we did not report it. Second, Goetzman and Jorion (1995) show that pre-1926 in-sample predictability is marginal, and our model suggests higher predictability in their period. Again, we

[19]...and perhaps even when the D/P coefficient is insignificant. Of course, such an investor should also use any other variable, perhaps including sunspots, with positive in-sample coefficients, as long as his theoretical prior does not put exactly zero weight on such variables. The priors obtained from theory here are unfortunately only heuristic: there is no firm reason why expected returns cannot follow a random walk or, with an appropriately strange process on the growth rates of dividends, even move in the opposite direction as the D/P ratio.

[20]We would also advocate that the profession should apply even more stringent criteria than "simple statistical significance in certain specifications" to believe in forecastability of the equity premium (and with enough search, some will surely turn up again). After all, forecasts of the equity premia are among the most widely researched subject areas in finance, and one should be cautious of any variable that works only in certain specifications or time-periods.

need a better theory as to how the relation between the dividend yield and future equity premia varies over time. Third, multi-year forecasts with our model are not sensible. As t goes to infinity, the dividend-yield coefficient (slowly) goes to negative infinity, the time intercept goes to positive infinity. Theoretical priors should force us to impose a minimum dividend coefficient of zero, and this constraint is not used in our specification. The model could be reestimated each year; or the zero constraint could be casually imposed, e.g., by forcing an inequality on the coefficients themselves.

Our own conclusion is that there are good indications that dividend-yield predictive coefficients are not stable but time-varying (and probably declining, but not necessarily in the simple deterministic log-time fashion used above), and that more investigation of the time-series properties of predictive coefficients is warranted. To this goal, our next section provides some additional descriptive information about the time-series of dividend yield coefficients.

5 A Description of Time-Varying In-Sample Coefficients

We did not have a formal theory of how the market learns: the log-time specification was arbitrary. It is thus useful to entertain some descriptive investigations into the time-series properties of the dividend-yield coefficient.

5.1 Time-Varying Coefficients

[Insert Table 7: Subperiod Regressions Predicting the Equity Premium (EQP)]

Table 7 partitions our sample into three subperiods of roughly equal length: the pre-war years, the post-war pre-1970 years, and the post-1970 years. In addition to the standard regression output, the final two columns contain the RMSE for the conditional and the unconditional model. Further, Figure 4 plots the relation between the equity premium and the lagged dividend yield.

[Insert Figure 4: The Equity Premium vs. Lagged Dividend Yield]

The subperiod coefficients further reveal some of the problems of the dividend yield predictor. Since 1970, there has been no apparent systematic relationship between the lagged dividend yield (measured either way) and the equity premium. There was an economically significant decline in the dividend yield coefficient (see also Figure 4) over time. This dividend yield coefficient decline explains why the unconditional model managed to outperform the conditional dividend yield model and why our "changing market" model outperforms both.

The final two columns provide the out-of-sample RMSE's in the subperiods. It is apparent that the problems of the dividend yield model are not isolated to the 1990's. Even in the 1946–1970 period, the conditional model fails to outperform the unconditional historical average. The dividend yield model just fails to predict throughout, using our out-of-sample periods.

It is interesting to note that *changes* in the dividend yield were relatively more successful in correlating with the future equity premium in the annual regressions in the last twenty years. For the difference in $D(t-1)/P(t-2)$, the coefficient is more stable; for the difference in $D(t-1)/P(t-1)$, the coefficient even increased over time. The fact that innovations in dividend yields may be more stable and thus better predictors than dividend yield levels further hints at a model in which it is "news" of dividend-yield surprises that matter, rather than a model in which everyone knows that the level of dividends calibrates expected returns.

5.2 Recursive Coefficients

[Insert Figure 5: Recursive Coefficient Estimates Predicting the Equity Premium With The Dividend Yield]

Another method to display model (in-)stability is to graph the time-series of the recursive (rolling) dividend yield coefficients used to predict out-of-sample. Each year, we reestimate the regression, and plot the coefficient and standard error in figure 5. The figure shows that the coefficient estimates are rather erratic during the great depression (which happens soon after our sample begins), but declined at

a mild and roughly constant rate (at a statistically insignificant but positive level) after 1936. The annual coefficient estimates have been statistically significant for quite a while, but they have also declined until about 1965.

5.3 The Optimal Coefficient Out-Of-Sample

What coefficient variations produces better out-of-sample prediction? That is, what kind of dividend-yield coefficient does it take to add useful information to the historical mean in terms of out-of-sample prediction?

[Insert Figure 6: The Perfect Dividend-Yield Coefficient To Maximize Relative Out-Of-Sample Performance]

Figure 6 plots the dividend-yield coefficient that perfectly fits the next out-of-sample data point, using as intercept the historical average equity premium up to each date.[21] Although some of the troughs and peaks necessarily line up with the well-known stock market ups and downs, the two are not the same (due to time-series changes in the dividend yield and changes in the historical equity premium mean). We also overlay a 5-year moving average version over the coefficient series; given that we believe that changes in the expected return/dividend-yield relationship are likely to be

Relative to its own historical average, the dividend yield coefficient steadily declined from about 1935 to the mid-1970's, reaching negative levels around 1960. This is similar to the conclusion reached in Pesaran and Timmermann (1995). (The important unanswered question remains of what differed during these periods.) However, we find only a sharp rise back to a zero relationship between yields and equity premia after 1975, not to a positive level that investors may have found useful.

[21] An alternative exercise would be to subtract from the dividend yield its own historical average. Unfortunately, the denominator often is close to zero, which explodes the coefficients.

6 Conclusion

Our paper has produced the following findings:

1. Good in-sample performance is no guarantee of out-of-sample performance in the equity premium prediction context. The simple dividend-yield predictions over the 1946–1997 period cannot beat the unconditional historical average equity premium on average, much less do so in a statistically significant manner. A naive market-timing trader who just assumed that the equity premium was "like it has been" would typically have outperformed a trader who employed dividend yield forecasting regressions.
2. *Any* remaining explanatory power of the dividend-yield out-of-sample in the post-war period is due to two years only, 1973 and 1974.
3. The dividend-yield coefficient is unstable. We could reject the hypothesis that the out-of-sample performance of $D(t-1)/P(t-2)$ is generated by a stable model.
4. A model in which the dividend-yield coefficient $D(t-1)/P(t-2)$ declines deterministically every year triples the in-sample adjusted r-square. It is—to our knowledge—the best in-sample forecaster of equity premia known today. The inference holds if the data ends in 1990.
5. A model in which the dividend-yield coefficient $D(t-1)/P(t-1)$ declines deterministically every year doubles the in-sample adjusted r-square. This model also produces the best out-of-sample forecasts of any model known to us, offers an annual forecast of 3-4% as of 1999, and indicates no marginal predictive ability of dividend-yields as of 1999.
6. We interprete the declining dividend-yield coefficient as indicative of a "changing market" hypothesis. It is consistent with an attempt of investors to "time the market" which in turn changed the equilibrium relation, and/or with an increase in the serial autocorrelation of the dividend-yield process itself, as documented in this paper.

Quo Vadis? Empirically, we need more investigation of the coefficient processes. Our log-decay specification, though highly successful, exceedingly simple and in line with a heuristic explanation, was a somewhat arbitrary process. One could allow for more extreme changes in the dividend-yield (e.g., structural breaks; see Pastor and Stambaugh (1998) andViceira (1997) for evidence of structural breaks in the equity premium itself), for stochastic changes in the coefficient, and/or for relatively weaker relationships with equity premia in recent years. Indeed, we feel more confident that the dividend-yield coefficient is changing than we feel confident arguing that it is declining in the log-time fashion. An additional route for investigation concerns a better modeling of dividend yield process itself. For example, one could improve the modeling of *surprises* rather than just dividend-yield levels. This is in the spirit of a model in which investors attempt to learn about and take advantage of any indicator of higher future returns, rather than of a model in which investors are aware that dividend yield levels simply calibrate time-changing expected return variation. For example, one could improve on our first steps in the estimation of a link between the persistence of the dividend-yield process and the dividend-yield's predictive ability.

Theoretically, it is not yet clear what theoretical models are consistent with the unit-root observed in dividend yields. Understanding the set of consistent processes could help one determine to what extent predictability process changes can be attributed to slow learning vs. efficient market processes. In general, we need more explicit modelling of the process by which the market learns, rather than the simple deterministic log-time ad-hoc assumption used in our paper. Such models would be particularly valuable if they offered explicit *speed*-of-adjustment prediction. Alternatively, the profession could benefit from theories of how other variables could influence the dividend yield coefficient. The ultimate goal would be to produce portfolio selection papers that can handle a more general dividend-premium relationship. Current models of portfolio selection, e.g., as in Barberis (1998), are inconsistent with the observed dividend process. By chasing a moving target as if it were fixed, the optimal investor updating process on the dividend-yield coefficient (and thus the investment profiles) is incorrect. With an extension of these partial equilibrium models to a moving target, one could use such models as the de-

mand component necessary to develop a full-equilibrium dividend-yield coefficient process.

References

Ball, R., 1978, "Anomalies in relationships between securities' yields and yield surrogates," *Journal of Financial Economics*, 6, 103-126.

Barberis, N., 1998, "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, ?, forthcoming.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria To Select Return Forecasting Models: What do we learn?," *Review of Financial Studies*, 12(2), 405-428.

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

Brennan, M. J., E. S. Schwartz, and R. Lagnado, 1997, "Strategic Asset Allocation," *Journal of Economic Dynamics and Control*, 21, 1377-1403.

Campbell, J. Y., and R. J. Shiller, 1988, "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1, 195-228.

Campbell, J. Y., and L. Viceira, 1998, "Consumption and Portfolio Decisions When Expected Returns are Time Varying," *Quarterly Journal of Economics*, ?, forthcoming.

Cochrane, J. H., 1997, "Where is the market going? Uncertain facts and novel theories," *Economic Perspectives*, 21, 3-37.

Dickey, D. A., and W. A. Fuller, 1979, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-31.

Dow, C. H., 1920, "Scientific Stock Speculation," *The Magazine of Wall Street*.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22(1), 3-25.

———, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25(1), 23-49.

Goetzman, W. N., and P. Jorion, 1995, "A Longer Look at Dividend Yields," *Journal of Business*, 68(4), 483-508.

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

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

Jarque, C. M., and A. K. Bera, 1987, "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163-172.

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

Lee, C., J. Myers, and B. Swaminathan, 1999, "What is the Intrinsic Value of the Dow?," *The Journal of Finance*, 54, forthcoming.

Lee, C. M., and B. Swaminathan, 1998, "Valuing the Dow: A Bottom-up Approach," working paper, Cornell University.

Liu, J., 1999, "Portfolio Selection in Stochastic Environments," working paper, Stanford University.

Lynch, A. W., 1999, "Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability," working paper, Stern School at New York University.

Nelson, C. R., and M. J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *The Journal of Finance*, 48(2), 641-661.

Pastor, L., and R. Stambaugh, 1998, "The Equity Premium and Structural Breaks," working paper, The Wharton School.

Pesaran, M., and A. Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *The Journal of Finance*, 50(4).

Rozeff, M., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management*, pp. 68-75.

Shiller, R. J., 1984, "Stock prices and social dynamics," *Brookings Papers on Economic Activity*, 2, 457-498.

Stambaugh, R., 1999, "Predictive Regressions," working paper, The Wharton School.

Viceira, L. M., 1997, "Testing for structural change in the predictability of asset returns," working paper, Harvard University.

Xia, Y., 1999, "Learning about Predictability: The Effect of Parameter Uncertainty on Dynamic Asset Allocation," working paper, UCLA.

Yan, S., 1999, "Can Dividend Yield Predict Future Returns: A Re-examination of Some Early Evidence," working paper, UCLA.

Table 1: Data Sources

Name	Description	Source	Availability
VWR	Value weighted stock return (incl. dividends)	CRSP	1926–1997
VWRX	Value weighted stock return (excl. dividends)	CRSP	1926–1997
TB	Treasury Bill Yield	IBBOTSON	1926–1997
EQP	The Equity Premium	VWR-TB	1926–1997
D(t-1)/P(t-2)	The dividend yield	VWR − VWRX	1926–1997
D(t-1)/P(t-1)	The dividend yield	$D(t-1)/P(t-2) \times [P(t-1)/P(t)]$	1926–1997

Explanation: Parenthesized expressions denote timing. When omitted, assume a time subscript of zero. In all regressions that follow, EQP will lead its predictors by one period. For example, the 1987 dividend yield (e.g., $D(1986 \text{ to } 1987)/P(1986)$) would be used to forecast 1988 equity premia EQP(1987 to 1988). $P(t)$ is stock market capitalization, used to convert $D(t)/P(t-1)$ into $D(t)/P(t)$, and was adjusted to account for the increase in market capitalization in July 1962 and December 1972, when the AMEX and NASDAQ came on-line.

Table 2: Descriptive Statistics

Entire Sample Period, 1926–1997

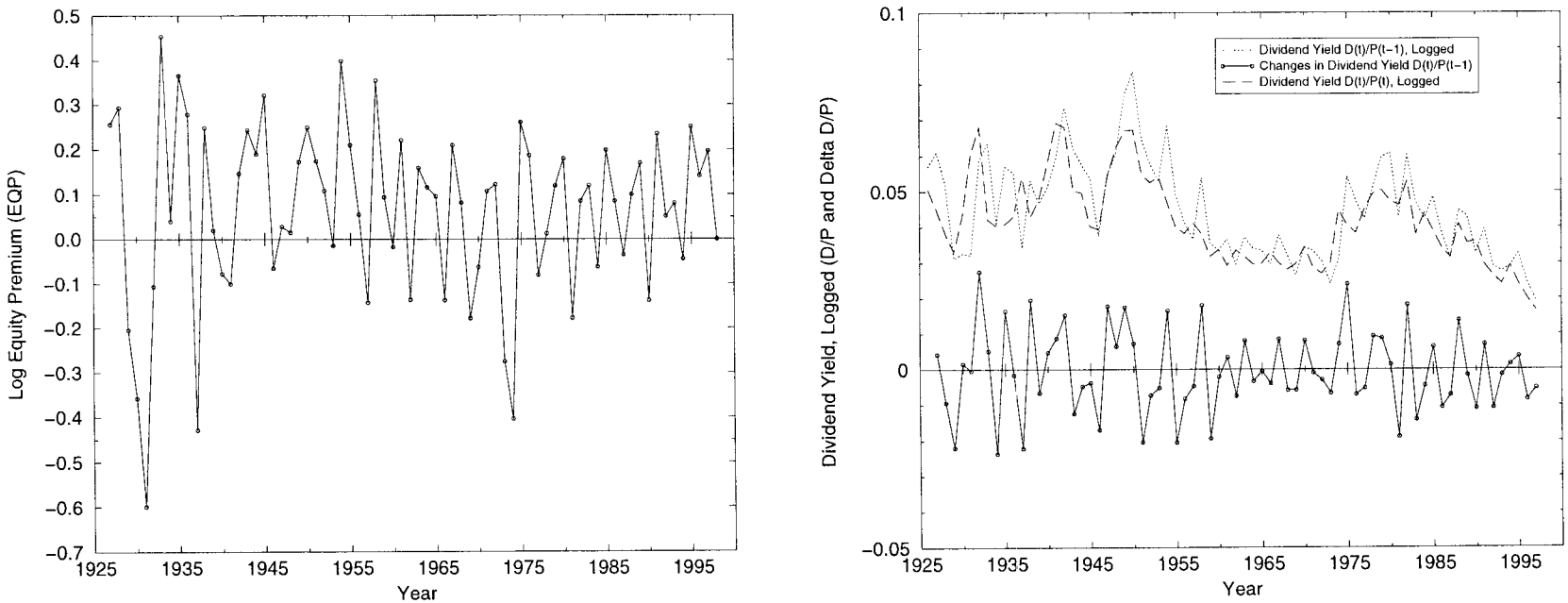
	In Levels				In Logs								
	N	**Mean**	**Sdev.**	**Median**	**Mean**	**Sdev.**	**Median**	**Min**	**Max**	**Skew**	**Kurt**	**JqBr**	**ADF**
VWR	72	12.5	20.4	15.0	10.0	19.6	14.0	−58.7	45.7	−1.0	4.3	17.0	−5.3
VWRX	72	7.9	19.7	10.9	5.8	19.9	10.4	−64.8	41.4	−1.0	4.4	18.4	−5.2
TB	72	3.8	3.2	3.2	3.7	3.1	3.2	0.0	13.7	0.8	3.4	9.0	−1.4
EQP	72	8.7	20.7	10.3	6.3	19.9	9.5	−59.8	45.4	−0.8	4.0	11.2	−5.3
$D(t)/P(t-1)$	72	4.6	1.5	4.4	4.5	1.4	4.3	2.1	8.3	0.5	2.5	3.4	−2.1
$D(t)/P(t)$	72	4.2	1.3	4.1	4.1	1.2	4.0	1.7	6.9	0.5	2.8	3.1	−1.9

Out-of-Sample Period, 1946–1997

	In Levels				In Logs								
	N	**Mean**	**Sdev.**	**Median**	**Mean**	**Sdev.**	**Median**	**Min**	**Max**	**Skew**	**Kurt**	**JqBr**	**ADF**
VWR	52	13.4	16.6	15.0	11.4	15.2	14.0	−32.7	40.7	−0.5	3.1	2.5	−5.4
VWRX	52	9.0	16.1	11.3	7.5	15.3	10.7	−37.1	35.9	−0.5	3.0	2.5	−5.5
TB	52	4.9	3.1	4.6	4.7	2.9	4.5	0.3	13.7	0.7	3.3	4.9	−1.8
EQP	52	8.5	17.2	10.3	6.7	15.8	9.5	−40.4	39.9	−0.6	3.4	3.3	−4.9
$D(t)/P(t-1)$	52	4.3	1.5	3.9	4.2	1.4	3.8	2.1	8.3	0.9	3.4	7.4	−1.6
$D(t)/P(t)$	52	3.9	1.2	3.8	3.8	1.1	3.7	1.7	6.7	0.7	3.0	3.8	−1.1

Explanation: All series are described in Table 1. Throughout the paper, they are measured in percent. Except where otherwise indicated, the paper reports only results using log variables. Log always means the natural log of 1 plus the value. Every mean and median is significantly different from zero at the 1% level. JqBr is the Jarque-Bera (Jarque and Bera (1987)) test for normality. The critical level to reject normality is 5.99 at the 95% level, 9.21 at the 99% level. ADF is the *Augmented Dickey-Fuller* (Dickey and Fuller (1979)) test for the absence of a unit root. An ADF value of -3.5 rejects the presence of a unit root at the 1% level (-2.9 at the 5% level; -2.6% at the 10% level).

Figure 1: Time Series Graphs



Explanation: The left graph plots the time series of the log equity premium (EQP). The right graph plots the dividend yield and changes in the dividend yield.

Table 3: Bivariate Regressions Predicting the Equity Premium (EQP) In-Sample

Dividend Yield is	CONST	Dividend Yield	R^2	$\overline{R}^2$	s.e.	N
$\frac{D(t-1)}{P(t-2)}$	−11.545	3.949	7.3	5.9	19.4	71
	-1.44	2.33				
	-1.23	2.21				
$\frac{D(t-1)}{P(t-1)}$	−14.443	4.977	8.6	7.2	19.3	71
	-1.70	2.54				
	-1.83	3.01				
$\left[\frac{D(t-1)}{P(t-2)}\right] - \left[\frac{D(t-2)}{P(t-3)}\right]$	6.261	3.828	5.0	3.6	19.7	70
	2.66	1.90				
	2.77	2.20				
$\left[\frac{D(t-1)}{P(t-1)}\right] - \left[\frac{D(t-2)}{P(t-2)}\right]$	6.131	1.688	0.4	−1.0	20.1	70
	2.54	0.55				
	2.53	0.77				

Explanation: Variables are described in Table 1, their descriptive statistics are in Table 2. The dependent variable, the (log) equity premium at year t (in percent), leads the independent variables by one year. The first row of each regression model is the coefficient, the second line its OLS t-statistic, the third line its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. The standard error (s.e.), R^2 and $\overline{R}^2$ (adjusted R^2) are quoted in percent.

Table 4: Properties of Forecast Errors Predicting The Equity Premium (EQP) Out-of-Sample

$D(t-1)/P(t-2)$ as Forecaster

	in-sample (1926-97)		in-sample (1946-97)		out-of-sample (1946-97)	
	DV	**UNC**	**DV**	**UNC**	**DV**	**UNC**
Mean	0.0	0.0	1.6	0.4	3.5	0.4
Std	19.3	19.9	15.5	15.8	16.9	16.0
RMSE[1]	19.3	19.9	15.6	15.8	17.3[1]	16.0[1]
Median	2.8	3.2	3.2	3.2	2.9	2.9
MAD	14.9	15.3	12.2	12.5	13.6	12.5
Max	33.8	39.1	32.6	33.5	44.3	33.7
Min	−61.1	−66.1	−38.3	−46.7	−33.7	−46.6

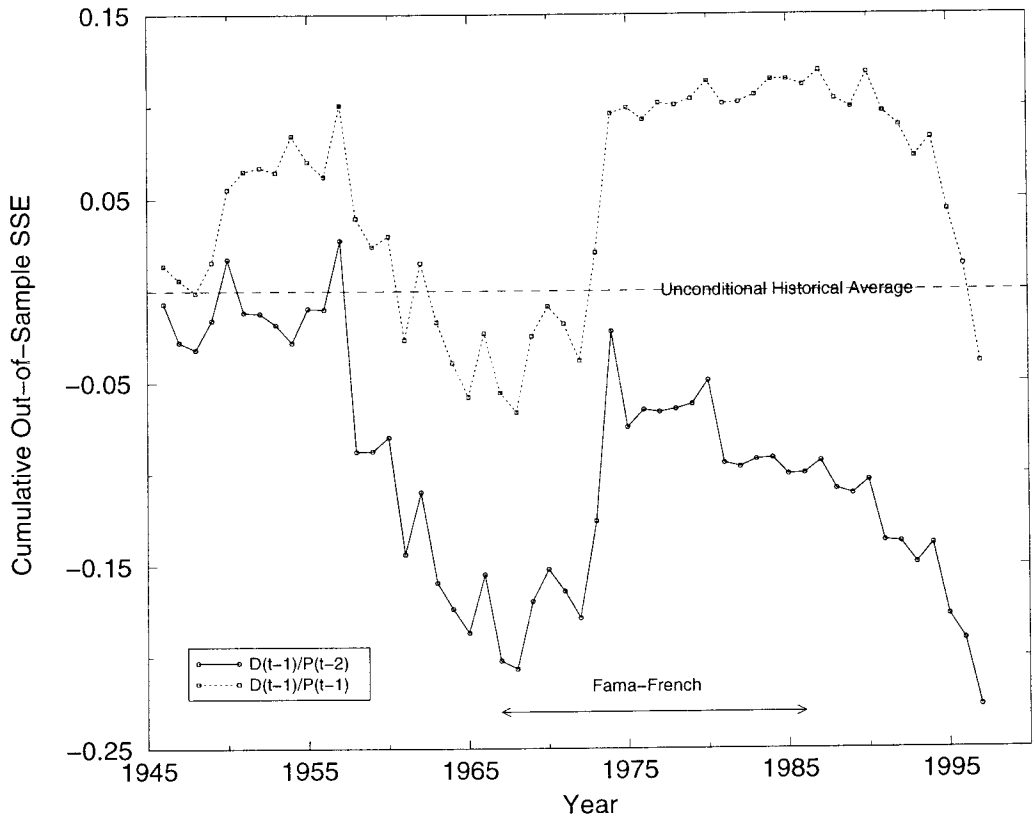
$D(t-1)/P(t-1)$ as Forecaster

	in-sample (1926-97)		in-sample (1946-97)		out-of-sample (1946-97)	
	DV	**UNC**	**DV**	**UNC**	**DV**	**UNC**
Mean	0.0	0.0	2.1	0.4	5.3	0.4
Std	19.2	19.9	14.9	15.8	15.3	16.0
RMSE[2]	19.2	19.9	15.1	15.8	16.2[2]	16.0[2]
Median	1.8	3.2	3.6	3.2	7.0	2.9
MAD	14.8	15.3	12.0	12.5	13.0	12.5
Max	31.0	39.1	29.4	33.5	37.8	33.7
Min	−66.8	−66.1	−40.8	−46.7	−37.6	−46.6

[1]: Jackknife(1.3)= 82.5%
[2]: Jackknife(0.2)= 57.7%

Explanation: This table describes the univariate properties of log equity premium prediction errors from one model that conditions on lagged log dividend yield (**DV**), and another model that uses only the historical average log equity premium as a forecast (**UNC**). The first two data columns compare the residuals of a single large regression estimated with data from 1926–1997 with the overall unconditional average over the 1926–1997 period. The middle two data columns partition these residuals to display only the residual prediction errors in the 1946–1997 period. The final two columns use only historical data in the estimation of the models to produce each forecast. Thus, each year the dividend yield regression and unconditional mean are reestimated with data available up to date in order to obtain an equity premium forecast (and forecast error). The Jackknife statistics are CDF's described in the text, and test for a statistically significant better prediction of either model (using the RMSE difference metric). They denote the likelihood that an RMSE difference less than the observed difference appears under the null hypothesis of equal performance. A value close to 1 indicates that the unconditional model outperforms with statistical significance, a value close to 0 indicates that the dividend model outperforms with statistical significance, a value close to 1/2 indicates that there is no statistical difference between the two models.

Figure 2: Cumulative Out-of-Sample Relative Out-of-Sample Sum-Squared Error Performance

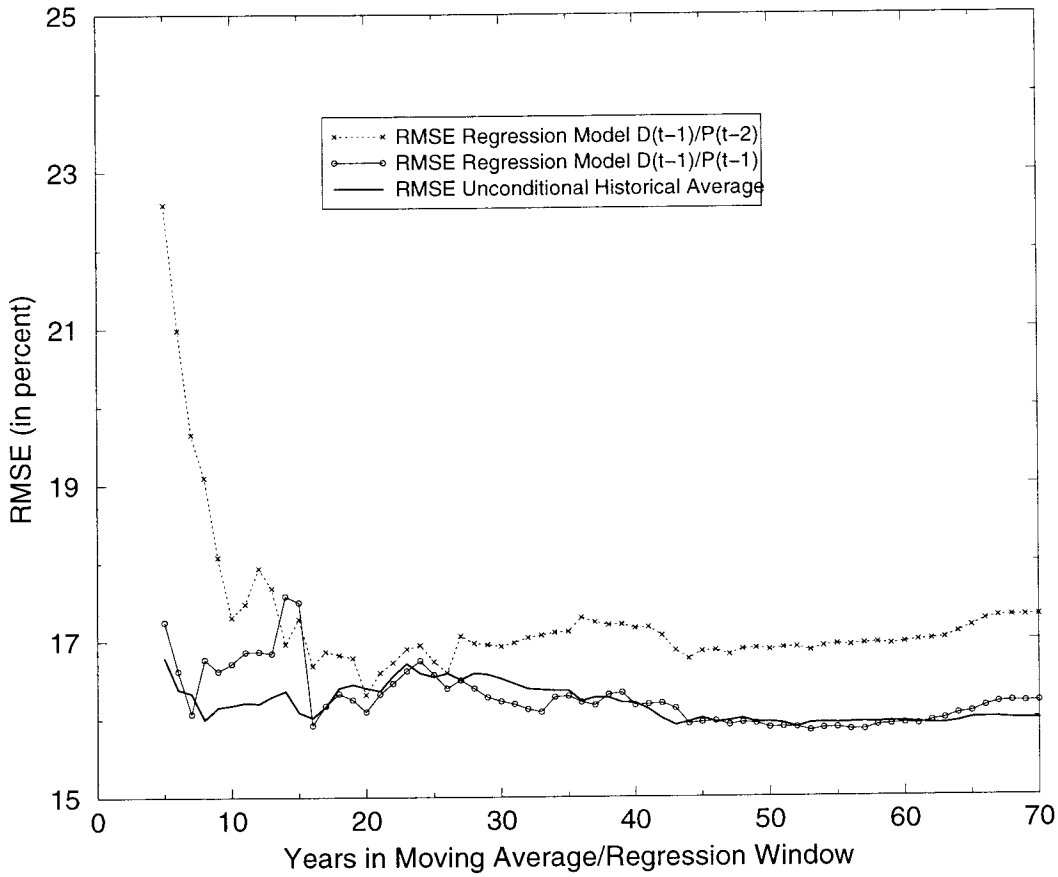


Explanation: This figure plots

$$f(T) = \sum_{t=1946}^{T} \mathrm{SE}_t^{\mathrm{Uncond}} - \mathrm{SE}_t^{\mathrm{Divyld}}$$

where SE_t is the squared out-of-sample prediction error in year t. The "uncond" SE is obtained when the historical up-to-date equity premium average is used to forecast the following year's equity premium. The "Divyld" SE's are obtained from rolling regressions with either $D(t-1)/P(t-2)$ or $D(t-1)/P(t-1)$ being the sole predictor of the following year's equity premium. For years where the slope of these lines is positive, the dividend-yield regression model predicts better than the unconditional average out-of-sample. This was the case, e.g., in 1973 and 1974.

Figure 3: Different Lengths of Estimation Windows To Predict the Equity Premium (EQP) Out-Of-Sample



Explanation: This figure plots the out-of-sample RMSE, forecasting the equity premium (since 1946) when the model estimation window before each forecast is restricted to a given number of years. (For the unconditional model, this is the simple moving average.) When the window requires more years than are available, the estimation window is shortened to begin with the first observation. The reported RMSE's in Table 4 are the levels at the right edge of each figure. They obtain when the estimation window is set to be "since inception."

Table 5: Time-Varying Dividend-Yield Coefficients

		$f(t)=$	CONST	DivYield	$f(t)$	DivYield $\times f(t)$	R^2	$\overline{R}^2$	s.e.	RMSE (unc=15.97)	Forecast (div=1.5%)	Forecast (div=1.2%)
Dividend Yield is D(t-1)/P(t-2)	M1	n/a	−11.545	3.949			7.27	5.93	19.44	17.31	−5.67	−6.83
			-1.23	2.21								
		t	−22.990	5.095	0.179		10.05	7.40	19.28	17.53	−2.36	−3.87
			-1.36	2.12	1.07							
	M2	t	−55.036	11.985	1.134	−0.222	16.86	13.14	18.68	18.06	21.42	22.68
			-2.17	2.67	2.23	-2.30						
		$\log(t)$	−30.936	4.773	4.697		10.92	8.26	19.20	17.01	−3.68	−5.09
			-1.53	2.29	1.31							
	M3	$\log(t)$	−116.283	22.906	30.337	−5.587	20.29	16.66	18.30	17.06	12.29	12.60
			-3.04	3.36	2.95	-2.97						
Dividend Yield is D(t-1)/P(t-1)	M1	n/a	−14.443	4.977			8.57	7.24	19.30	16.20	−7.03	−8.51
			-1.83	3.01								
		t	−33.310	7.348	0.256		13.60	11.06	18.90	15.83	−3.65	−5.82
			-2.17	3.41	1.45							
	M2	t	−51.779	11.607	0.810	−0.140	15.75	11.97	18.80	16.54	9.41	9.00
			-2.89	3.78	2.31	-1.86						
		$\log(t)$	−40.542	6.590	5.825		14.07	11.51	18.86	15.53	−5.74	−7.69
			-2.04	3.46	1.46							
	M3	$\log(t)$	−93.784	19.043	21.865	−3.868	16.61	12.82	18.72	15.55	3.67	2.95
			-3.26	3.14	2.56	-1.93						

Explanation: Variables are described in Table 1, except t (a year index that subtracts 1926), their descriptive statistics are in Table 2. The dependent variable, the equity premium, leads the independent variables by one year. The first row of each regression model is the coefficient, the second line its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. The RMSE column gives the out-of-sample model RMSE of the regression models, each estimated in a rolling fashion (only with data up-to-date). The final two columns provide an annual return forecast (in percent) for 1999, given a dividend yield of either 1.5% or 1.2%.

Table 6: Time-Varying Dividend-Yield Coefficients as of 1990

Dividend Yield is D(t-1)/P(t-2)

	CONST	Dividend Yield	log(t)	Dividend Yield × log(t)	R^2	$\overline{R}^2$	s.e.	RMSE (unc=16.45)	RMSEDiff	Jackknife
M1	−17.533	5.012			10.58	9.16	19.69	17.15	+0.70%	67%
	-1.75	2.67								
M3	−111.357	21.962	28.462	−5.226	20.17	16.18	18.93	17.77	+1.32%	97%
	-2.57	2.84	2.31	2.32						

Dividend Yield is D(t-1)/P(t-1)

	CONST	Dividend Yield	log(t)	Dividend Yield × log(t)	R^2	$\overline{R}^2$	s.e.	RMSE (unc=16.45)	RMSEDiff	Jackknife
M1	−23.575	6.806			13.46	12.09	19.37	15.60	-0.84%	25%
	-3.18	4.47								
M3	−71.213	14.040	13.486	−1.998	17.95	13.85	19.19	15.70	-0.75%	7%
	-1.91	1.85	1.20	-0.80						

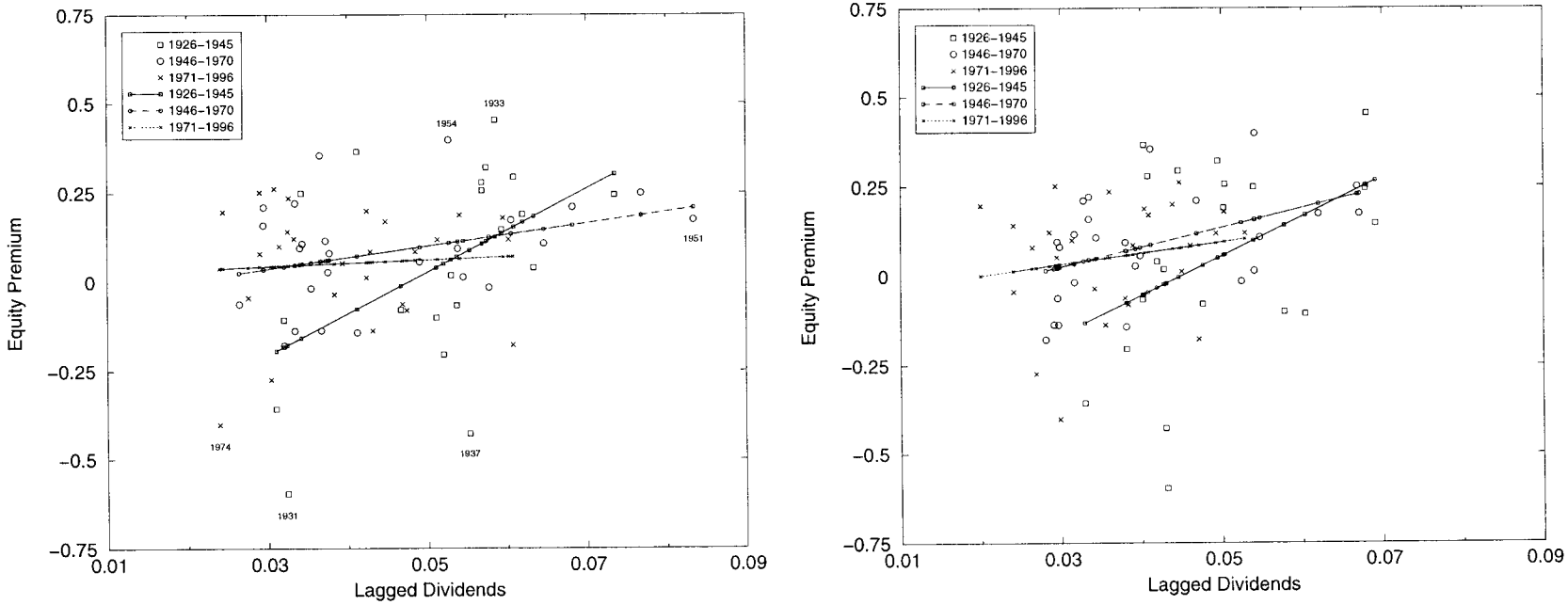
Explanation: Variables are described in Table 1, except t (a year index that subtracts 1926), their descriptive statistics are in Table 2. The dependent variable, the equity premium, leads the independent variables by one year. The first row of each regression model is the coefficient, the second line its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. The RMSE column gives the out-of-sample model RMSE of the regression models, each estimated in a rolling fashion (only with data up-to-date). The Jackknife measures the probability of observing an RMSE difference as large in favor of the conditional model; a number close to 1 indicates that an observer should have been so disappointed with the conditional model that he should have abandoned it in favor of the unconditional model.

Table 7: Subperiod Regressions Predicting the Equity Premium (EQP)

	In-Sample Regression Statistics							Out-of-Sample	
	CONST	**DivYld**	R^2	$\overline{R}^2$	**s.e.**	**N**	**sd(EQP)**	**model**	**uncond**
			DivYld is $D(t-1)/P(t-2)$						
1926-45	−55.781	11.718	23.3	19.0	25.7	20	28.5		
	−1.73	2.06							
1946-70	−5.939	3.213	11.4	7.5	14.5	25	15.3	17.6	15.6
	−1.08	3.12							
1971-97	2.048	0.833	0.3	−3.8	17.0	16	16.4	17.4	16.5
	0.12	0.23							
			DivYld is $D(t-1)/P(t-1)$						
1926-45	−49.598	11.034	16.7	12.0	26.8	20	28.5		
	−1.89	2.47							
1946-70	−13.860	5.477	20.6	17.2	13.7	25	15.3	15.7	15.7
	−2.40	3.82							
1971-97	−5.453	2.964	2.6	−1.4	16.8	26	16.4	16.9	16.5
	−0.30	0.70							
			DivYld is $D(t-1)/P(t-2) - D(t-2)/P(t-3)$						
1926-45	3.623	4.335	4.5	−1.1	29.1	19	28.5		
	0.52	1.08							
1946-70	8.889	2.584	4.1	−0.1	15.0	20	15.3	15.7	15.6
	3.58	2.15							
1971-97	5.614	4.860	8.8	5.0	16.2	16	16.4	15.9	16.5
	2.04	2.01							
			DivYld is $D(t-1)/P(t-1) - D(t-2)/P(t-2)$						
1926-45	3.354	−2.417	0.9	−4.9	29.6	19	28.5		
	0.47	−0.76							
1946-70	8.806	5.608	4.6	0.5	15.0	25	15.3	16.1	15.7
	3.26	1.43							
1971-97	5.863	7.886	9.4	5.6	16.2	26	16.4	16.5	16.5
	1.89	3.28							

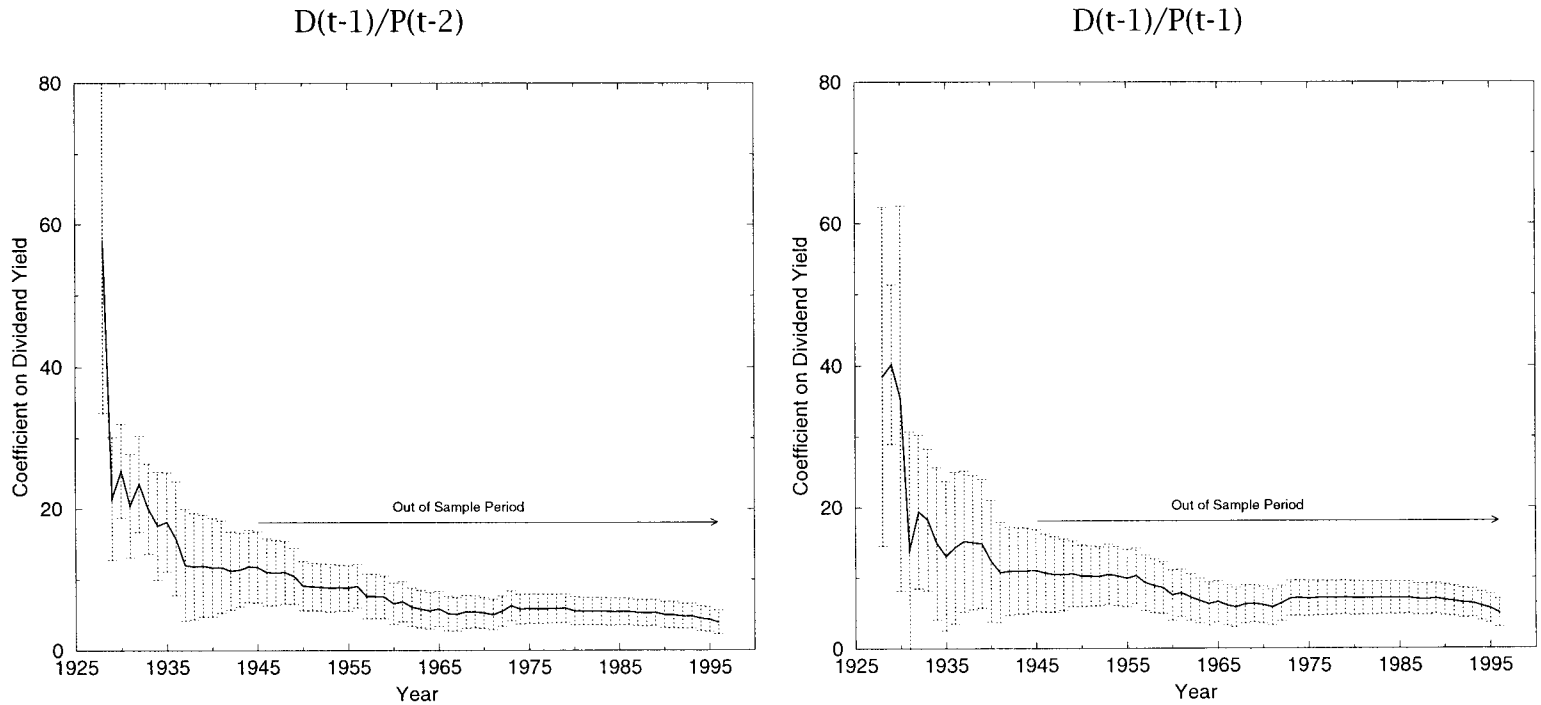
Explanation: Variables are described in Table 1, their descriptive statistics are in Table 2. The dependent variable, the equity premium, leads the independent variables (underlined header) by one year. The first row of each regression model is the coefficient, the second line its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. The out-of-sample model columns give the RMSE of both the regression models and the unconditional historical equity premium, each estimated with data up-to-date.

Figure 4: The Equity Premium vs. Lagged Dividend Yield



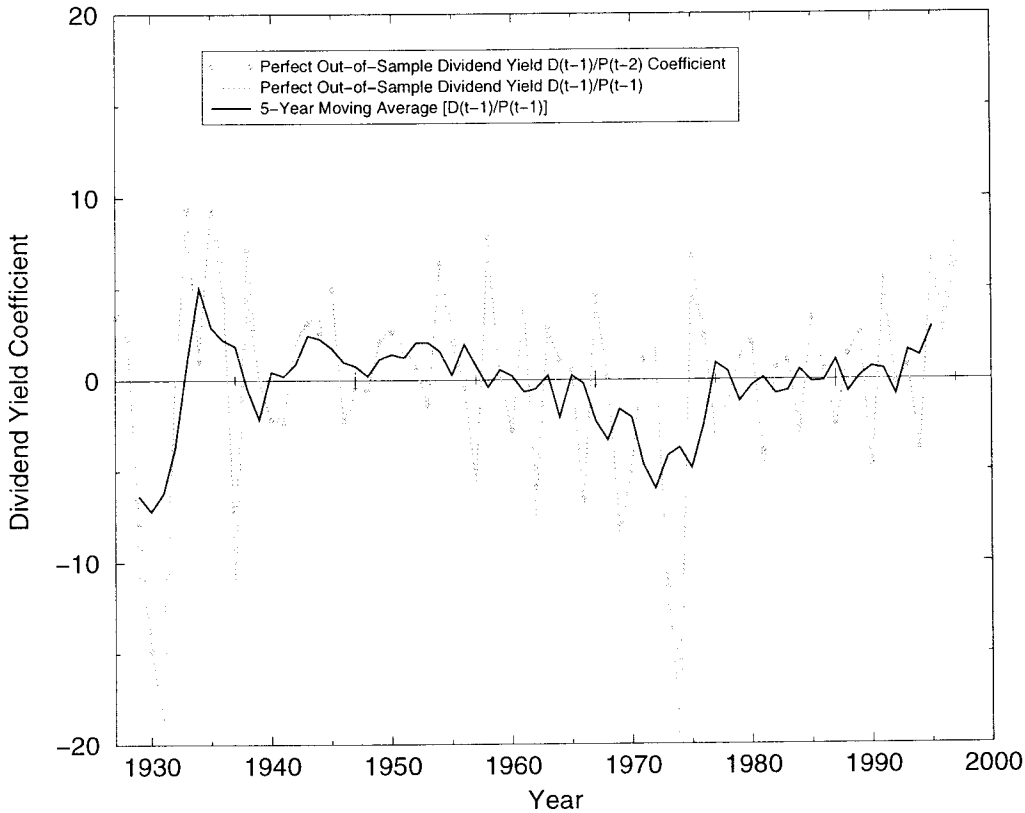
Explanation: Variables are described in Table 1, their descriptive statistics are in Table 2. The data frequency is annual. The dependent variable, the equity premium, leads the independent variable, the dividend yield, by one year. Some observations are marked to indicate the year of the dependent variable. The lines are fitted from the regressions.

Figure 5: Recursive Coefficient Estimates Predicting the Equity Premium With The Dividend Yield



Explanation: These figures plot the recursive coefficient estimates (i.e., using only historically available data at each point) in a regression predicting the (log) equity premium with the (log) dividend yield. The left graph plots the D(t-1)/P(t-2) coefficient, the right graph plots the D(t-1)/P(t-1) coefficients, both obtained from bivariate regressions. The bars denote plus and minus *one* standard deviation.

Figure 6: The Perfect Dividend-Yield Coefficient To Maximize Relative Out-Of-Sample Performance



Explanation: This figure plots the dividend-yield coefficient that performs best out-of-sample, i.e.,

$$\beta(t) = \frac{\text{EQP}(t) - \text{Avg}(\text{EQP}(j), j = 1926...t-1)}{DVY(t-1)}$$

The fat line plots the 5-year moving average on D(t-1)/P(t-1), but the equivalent D(t-1)/P(t-2) graph is visually indistinguishable.

Appendices

A The Relation Between The Persistence of The Dividend Yield and Its Predictive Ability

$$
\begin{aligned}
D_t &= \text{Dividends paid during } [t-1, t] \\
P_t &= \text{Price at end of period } t \\
R_t &= \text{Gross Return for period } [t-1, t] \\
d_t &= \text{log dividends during } [t-1, t] \\
p_t &= \text{log price at end of } t \\
r_t &= \text{log stock return for period } = \log(R_t) = \log\left(\frac{P_t + D_t}{P_{t-1}}\right) \\
\delta_t &= \text{log dividend price } \textbf{\textit{ratio}} \equiv d_t - p_t \equiv \log\left(\frac{D_t}{P_t}\right) \\
\theta_t &= \text{log dividend price } \textbf{\textit{yield}} \equiv d_t - p_{t-1} \equiv \log\left(\frac{D_t}{P_{t-1}}\right) \\
\Delta d_t &= \text{log dividend growth rate} \equiv d_t - d_{t-1} \equiv \log\left(\frac{D_t}{D_{t-1}}\right)
\end{aligned}
$$

Dividend Price *Ratio* D_t/P_t

$$
\begin{aligned}
r_{t+1} &= \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \\
&= \log\left(\frac{P_{t+1} + D_{t+1}}{P_{t+1}} \cdot \frac{P_{t+1}}{P_t}\right) = \log\left(\frac{P_{t+1}}{P_t}\right) + \log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) = p_{t+1} - p_t + \log\left(1 + e^{\delta_{t+1}}\right) \\
&= \log\left(\frac{P_{t+1} + D_{t+1}}{D_{t+1}} \cdot \frac{D_t}{P_t} \cdot \frac{D_{t+1}}{D_t}\right) = \log\left(\frac{P_{t+1}}{D_{t+1}} \cdot \frac{D_t}{P_t} + \frac{D_t}{P_t}\right) + \log\left(\frac{D_{t+1}}{D_t}\right) = \log\left(e^{\delta_t - \delta_{t+1}} + e^{\delta_t}\right) + \Delta d_{t+1}
\end{aligned}
$$

Assume $\{\delta_t\}$ follows a stationary process with mean $\bar{\delta} = \overline{d - p}$. Expand $f(x) = \log(1 + e^x)$ using Taylor expansion around $\bar{x}$.

$$
\begin{aligned}
f(x_{t+1}) &\approx f(\bar{x}) + f'(\bar{x})(x_{t+1} - \bar{x}) \\
\log\left(1 + e^{\delta_{t+1}}\right) &\approx \log\left(1 + e^{\bar{\delta}}\right) + \frac{e^{\bar{\delta}}}{1 + e^{\bar{\delta}}}\left(\delta_{t+1} - \bar{\delta}\right)
\end{aligned}
$$

Define $\kappa = 1 \Big/ \left(1 + e^{\bar{\delta}}\right) = 1 \Big/ \left(1 + \overline{D_t/P_t}\right)$ and $k = -\log\kappa - (1-\kappa)\log\left(\frac{1}{\kappa} - 1\right)$. After some algebra, we obtain

$$
r_{t+1} \approx \kappa p_{t+1} - p_t + (1 - \kappa) d_{t+1} + k
$$

$$= \quad -\kappa\delta_{t+1} + \delta_t + \Delta d_{t+1} + k \tag{23}$$

Taking covariance of eqn(1) with δ_t and dividing by variance of δ_t, we get

$$\begin{aligned} \frac{cov(r_{t+1}, \delta_t)}{var(\delta_t)} &= 1 - \kappa \cdot \frac{cov(\delta_{t+1}, \delta_t)}{var(\delta_t)} + \frac{cov(\Delta d_{t+1}, \delta_t)}{var(\delta_t)} \\ &or \\ \beta_{r_{t+1},\delta_t} &= 1 - \kappa \cdot \beta_{\delta_{t+1},\delta_t} + \beta_{\Delta d_{t+1},\delta_t} \end{aligned} \tag{24}$$

If δ_t follows an AR(1) process with persistence parameter β and the dividend growth rate is constant, we get $\beta_{r_{t+1},\delta_t} = 1 - \kappa \cdot \beta_{\delta_{t+1},\delta_t}$. Calibrating to US data, we have $\overline{D_t/P_t} \approx 4.2\%$. This implies that $\kappa \approx 0.96$ and $\boxed{\beta_{r_{t+1},\delta_t} \approx 1 - 0.96 \cdot \beta_{\delta_{t+1},\delta_t} + \beta_{\Delta d_{t+1},\delta_t}}$. It is straightforward to repeat this exercise for the dividend yield, $D(t)/P(t-1)$, rather than the dividend price ratio $D(t)/P(t)$ as long as $\beta_{\Delta d_{t+1},\delta_t} = 0$. The equivalent final expression is $\boxed{\beta_{r_{t+1},\delta_t} \approx \beta_{\delta_{t+1},\delta_t}(1 - 0.96 \cdot \beta_{\delta_{t+1},\delta_t})}$, where $\delta_t \equiv D(t)/P(t-1)$. This shows that when $\beta_{\delta_{t+1},\delta_t}$ is small enough, $\beta_{\delta_{t+1},\delta_t}$ can have a positive relationship with β_{r_{t+1},δ_t}. (This only appears to be the case in the first few years of our sample.)

B Fama and French (1988, 1989)

This appendix reconciles the $D(t-1)/P(t-2)$ dividend-yield model to the evidence in Fama and French (1988) (henceforth, FF). Fama and French also examine out-of-sample performance and come to different conclusions for $D(t-1)/P(t-2)$ for sample periods ending in the 1980's. (Our conclusions are more similar for $D(t-1)/P(t-1)$ models, again when the data ends in the 1980's; naturally, they could not consider the experience of the 1990's.) We here describe a number of small differences between the FF specification and our own, which turn out to be largely inconsequential.

1. FF use a different definition of the dividend yield, which ignores the within-year timing of dividends. Our definition of the dividend yield reinvests the dividends to adjust for the timing of annual dividend payments.[22] Also, we use log-dividend yields, not simple dividend yields.
2. FF use only NYSE firms, while we use all firms (NASDAQ, NYSE, AMEX).
3. FF focus on value-weighted and equal-weighted stock returns (both nominal and inflation-adjusted), while we focus on the equity premium. For value-weighted returns in particular, the use of nominal returns produces better out-of-sample performance than their inflation-adjusted returns. We focus on equity-premia, which are (mostly) invariant to inflation and less predictable (see Table 8 below).
4. FF use only a 30-year estimation window. In Figure 3, we showed that estimation windows of different sizes do not affect our inference.

[22] This choice is mostly a matter of taste on an annual horizon. On the multi-year horizon, however, differences in the definition of the dividend yield can become important. Our own measure would have reported only a 7% R^2 in the four-year regressions, in contrast to 13.6% obtained when the FF dividend yield definition is used (see their Table 3).

5. Although FF define an out-of-sample R^2 by dividing the MSE by the variance of the returns to be forecast, they do not consider an alternative ex-ante benchmark prediction.[23] De-facto, the normalization of the MSE with ex-post realized return variance implies a benchmark to which a trader would not have had access to. The R^2 metric is a useful normalization to report results across different dependent variables or periods, but it does not offer an implementable alternative—a trader would not have known the average *ex-post* realization of the equity premium *ex-ante*.

   Our own paper follows Fama and French (1989) in exploring a different naive benchmark, which uses historical *moving* averages of the (lagged) equity premium as out-of-sample predictions. Indeed, both the benchmark and the regression model commonly produce negative out-of-sample R^2 using the Fama and French (1988) definition. In the 1967–1986 period, using the FF definitions and sample, the unconditional forecast has an out-of-sample R^2 of about -10%. This is significantly worse than the FF out-of-sample R^2 of about 3%, which indicates to a reader that the conditional model is preferable. (Of course, a trader interested in making equity premium predictions might not have had access to *any* models which produced positive out-of-sample R^2's.)

[Insert Table 8: RMSE and Out-of-Sample R^2 By Definition and Sample Period]

Table 8 compares the effects of specification differences. The first row reproduces the out-of-sample performance in Fama and French (1988, Table 5), using their definitions, but adds the equity premium and RMSE's. The out-of-sample R^2's are virtually identical. The out-of-sample R^2 for equity premia hint at less forecastability. The next two rows describe the difference in RMSE and R^2 in forecast prediction errors when we apply the FF definitions (their dividend yield definition, NYSE firms only, 30-year estimation window) and our own definitions (compounding dividend yield definition, all firms, estimation window beginning in 1926) in the 1967–1986 period. The final two rows describe the difference in RMSE and R^2 in forecast prediction errors when we apply the FF definitions (their dividend yield, NYSE firms only) and our own definitions in the 1946–1997 period, using all available observations for estimation. Negative values imply that the conditional regression model outperforms the unconditional mean model.

The numbers show that differences in our specification, especially as far as the equity premium is concerned, are cosmetic. Our definitions tend towards zero, giving less favor to the conditional model in the 1967–1986 period and more favor to the conditional model in the 1946–1997 period. The table also shows that our conclusion of frequent superior performance by the unconditional mean does not apply to equal-weighted returns. Indeed, the -1.3% in the 1946–1997 sample using the FF definitions lies at the 7th percentile under the null hypothesis' distribution, almost enough to attribute superior performance to the regression model at conventional significance levels. (All other performance differences predicting the equal-weighted returns are insignificant.)

The reason why we find *better* performance of the unconditional model turns out to be the the choice of subperiod, as explained in the text and Figure 2. Still, the power of the FF tests in their 1967–1986 period did not rise to "reasonable" statistical reliability. Given that we have established that the in-sample performance in the equity premium context is typically significant even when it underperforms an unconditional model out-of-sample, the out-of-sample evidence must stand on its own in establishing superiority over the unconditional benchmark.

[23] To compare different variables or time periods, it makes more sense to use R^2. When the variable and out-of-sample period are the same, the sign of the RMSE difference is of course always the opposite of the sign of the out-of-sample R^2 difference.

C Other Prominent Equity Premium Forecasting Measures

C.1 The Dividend Payout Rate

[Insert Table 9: Using the Dividend/Earnings Payout Rate To Predict The Equity Premium (EQP)]

We also repeated the forecasting horserace with the dividend payout rate (dividends as a function of earnings), as proposed in Lamont (1998). We could only do this with quarterly and annual data, because monthly data is unavailable. We use Lamont's earnings measure, but use the dividend on the value-weighted return instead of the dividend yield on the S&P500. (Lamont also uses a small adjustment on the t-bill rate that we do not replicate.)

The regressions in Table 9 confirm Lamont's regressions: On a quarterly basis, the payout rate is statistically significant in-sample with a t-statistic of 2.3 in the 1947–1994 period. On an out-of-sample basis, using the Jackknife, the -0.07% difference in RMSE between the D/E_{-1} model and the unconditional history for the same sample period lies at the 20th percentile. Consequently, like the dividend yield itself, the dividend payout rate does not have statistical significance at conventional significance level. As Lamont points out, there is no significance in the annual regressions. On an out-of-sample basis, the annual RMSE difference of 0.90% (in favor of the unconditional forecast) lies at the 90th percentile under the null hypothesis.

C.2 A Value-Yield Measure

Lee, Myers, and Swaminathan (1999) construct a V/P measure, which uses a value measure based on accounting analysis as its numerator instead of the dividends. The authors were kind enough to make this data available to us.

[Insert Table 10: Using the Lee-Myers-Swaminathan V/P Measures To Predict the Equity Premium (EQP) and the Return on the Dow-Jones 30]

Table 10 repeats our forecasting exercise with their variables. Only monthly data are available. The first panel forecasts our equity premium, the second the Lee, Myers, and Swaminathan (1999) Dow-30 return. (The latter uses the DJ dividend yield, rather than our dividend yield.) The independent variable in the regressions are two of their best measures, VP3TB(3f) and VP12TB(3f), as indicated by Lee, Myers, and Swaminathan (1999, Table 8). For more details on variable construction, please refer to the original paper.

The left-hand side of the table confirms that their measures significantly outperform either the dividend yield or the unconditional mean forecast in-sample. The right-hand side of the table shows that their V/P measures outperform the D/P variable, but not the simple unconditional mean. In Panel A, VP3TB(3f) performs as well but no better than the unconditional mean. In Panel B, VP3TB(3f) beats the unconditional historical average by a statistically insignificant[24] 0.01% on the RMSE statistic, and underperforms on the MAD statistic. Further, VP3TB(3f) is significantly more biased. VP12TB(3f) is inferior to the unconditional historical average on all metrics.

[24]The 0.01% RMSE difference lies at the 55th percentile on the Jackknifed distribution.

Thus, although the Lee, Myers, and Swaminathan (1999) V/P measure indeed outperforms a D/P measure both in-sample and out-of-sample, it cannot reliably beat the simple unconditional historical average out-of-sample in our experiment. A market-timing trader would have performed just as well using the historical average as he would have been using a V/P measure. In its defense, unlike other more commonly used variables, the Lee, Myers, and Swaminathan (1999) V/P measures at least do not systematically underperform the unconditional mean.[25]

Lee and Swaminathan (1998) independently have run out-of-sample tests on their V/P measure by comparing a "static asset allocation" strategy to a dynamic (V/P based) asset allocation strategy. They find that it would take some rather large bets and fairly long holding periods to realize significant gains from their dynamic strategy, but they do find statistically significant results (before transaction costs) with holding periods of 24 months. There are other differences, such as their focus only on the post-1979 period, because it is the only period when analyst earning forecasts are available. (Pre-1979 V/P measures are weaker because they rely on time-series rather than analysts' forecasts.)

[25]Of course, compared with the D/P measure, the V/P measure was designed more recently. The jury as to its long-term predictive ability is still out.

Table 8: RMSE and Out-of-Sample R^2 By Definition and Sample Period

Specification		RMSE Metric					Out-of-Sample R^2 Metric				
		VWR	VWR-π	EWR	EWR-π	EQP	VWR	VWR-π	EWR	EWR-π	EQP
FF 1967–1986	**DV**	16.1	17.9	21.4	22.3	17.9	13.3	7.6	16.0	16.7	2.8
FF 1967–1986	**DV–UNC**	-1.4	-1.5	-2.5	-3.0	-1.2	-16.2	-16.5	-20.5	-24.0	-13.3
GW 1967–1986	**DV–UNC**	-1.2	-1.3	-1.1	-1.6	-1.1	-13.1	-14.4	-7.7	-11.2	-12.4
FF 1946–1997	**DV–UNC**	1.9	0.8	-1.3	-1.7	1.4	26.5	9.5	-13.5	-17.7	18.5
GW 1946–1997	**DV–UNC**	1.9	0.7	0.1	-0.8	1.3	26.4	8.4	0.8	-7.1	17.5

Explanation: VWR is the nominal value-weighted (annual) return; VWR $-\pi$ is the inflation-adjusted value-weighted return; EWR is the nominal equal-weighted (annual) return; EWR $-\pi$ is the inflation-adjusted equal-weighted return; EQP is the equity premium. The first row describes the out-of-sample performance predicting the variable in each column header with its respective annual lagged dividend yield. It reproduces the out-of-sample statistics in Fama and French (1988, Table 5), using their definitions, adding the equity premium and RMSE's. The next two rows describe the differences in the RMSE metric and the out-of-sample R^2 metric for forecast prediction errors from the conditional (**DV**) and unconditional (**UNC**) model, when we apply the FF definitions (their dividend yield definition, NYSE firms only, 30-year estimation window) and our own GW definitions in the 1967–1986 period. Negative values imply that the conditional regression model outperforms the unconditional mean model. The final two rows repeat this for the 1946–1997 period, using all available observations for estimation.

Table 9: Using the Dividend/Earnings Payout Rate To Predict The Equity Premium (EQP)

						In-Sample		Out-of-Sample		
Period	**CONST**	**D/E$_{-1}$**	R^2	$\overline{R}^2$	**N**	**s.e.**	**unc. std.dev.**	**RMSE period**	**D/E$_{-1}$ RMSE**	**unc RMSE**
						Quarterly				
1935:1 - 1994:4	5.348	0.055	2.1	1.7	240	8.52	8.61	1946 : 1 – 1994 : 4	7.94	7.99
	2.99	2.07								
1947:1 - 1994:4	6.718	0.074	3.4	2.9	192	7.78	7.89	1947 : 1 – 1994 : 4	7.84	7.91
	3.19	2.31								
						Annual				
1926 - 1997	8.528	0.039	0.2	−1.2	71	20.16	19.90	1946 – 1997	16.94	15.97
	1.34	0.39								
1947 - 1994	9.179	0.039	0.2	−2.0	48	16.27	15.89	1947 – 1994	17.01	16.11
	1.40	0.35								

Explanation: The data frequency is marked. The dependent variable, the equity premium, leads the independent variable, the dividend/earnings payout rate (D/E), as proposed in Lamont (1998), by one period. The first row is the coefficient, the second the Newey-West t-statistic. The in-sample regression s.e. is divided by $N - 2$, the unconditional std.dev. is divided by $N - 1$. The out-of-sample RMSE's reestimates the prediction model with data available up-to-date. The "unc" model uses the up-to-date historical average to predict the equity premium.

Table 10: Using the Lee-Myers-Swaminathan V/P Measures To Predict the Equity Premium (EQP) and the Return on the Dow-Jones 30

Panel A: Predicting the Equity Premium (EQP)

	in-sample (1979:01-1996:06)				out-of-sample (1980:01-1996:06)			
	VP3TB(3f)	**VP12TB(3f)**	**DIV**	**UNC**	**VP3TB(3f)**	**VP12TB(3f)**	**DIV**	**UNC**
Mean	0.00	0.00	0.00	0.00	0.40	0.13	−0.04	0.04
Std	4.26	4.27	4.35	4.34	4.38	4.43	4.43	4.40
RMSE	4.26	4.27	4.35	4.34	4.40	4.43	4.43	4.40
Median	0.56	0.53	0.34	0.42	0.77	0.50	0.36	0.45
MAD	3.18	3.16	3.15	3.14	3.28	3.32	3.16	3.16
Max	11.05	10.72	11.00	11.04	11.75	11.03	10.81	11.09
Min	−25.64	−26.11	−26.86	−26.69	−25.14	−25.45	−27.34	−26.82

Panel B: Predicting The Return on the Dow-Jones

	in-sample (1979:01-1996:06)				out-of-sample (1980:01-1996:06)			
	VP3TB(3f)	**VP12TB(3f)**	**DJDP**	**UNC**	**VP3TB(3f)**	**VP12TB(3f)**	**DJDP**	**UNC**
Mean	0.00	0.00	0.00	0.00	0.47	0.23	0.20	0.07
Std	4.31	4.32	4.41	4.41	4.42	4.47	4.56	4.46
RMSE	4.31	4.32	4.41	4.41	4.45	4.48	4.57	4.46
Median	−0.01	0.15	0.07	−0.01	0.44	0.26	0.21	−0.04
MAD	3.25	3.23	3.26	3.26	3.33	3.34	3.32	3.27
Max	12.48	12.14	12.54	12.49	13.08	12.42	12.42	12.45
Min	−24.48	−24.94	−25.45	−25.51	−24.25	−24.50	−26.45	−25.78

Explanation: This table predicts the monthly equity premium and the Dow-Jones return with two measures of V/P (value price ratio) obtained from Lee, Myers, and Swaminathan (1999), a dividend yield measure (our own in Panel A; the Lee, Myers, and Swaminathan (1999) Dow dividend yield in Panel B), and the unconditional mean. The left-hand side display statistics for the in-sample residuals. The right-hand side displays statistics for out-of-sample forecast errors, when only up-to-date historical information is used to estimate (rolling) regressions and historical means for purposes of forecasting.