A Comprehensive Look at The Empirical Performance of Equity Premium Prediction

Ivo Welch
Brown University Department of Economics NBER

Amit Goyal
Emory University Goizueta Business School

Our article comprehensively reexamines the performance of variables that have been suggested by the academic literature to be good predictors of the equity premium. We find that by and large, these models have predicted poorly both in-sample (IS) and out-of-sample (OOS) for 30 years now; these models seem unstable, as diagnosed by their out-of-sample predictions and other statistics; and these models would not have helped an investor with access only to available information to profitably time the market. (*JEL* G12, G14)

Attempts to predict stock market returns or the equity premium have a long tradition in finance. As early as 1920, Dow (1920) explored the role of dividend ratios. A typical specification regresses an independent lagged predictor on the stock market rate of return or, as we shall do, on the equity premium,

$$\text{Equity Premium}(t) \;=\; \gamma_0 \;+\; \gamma_1 \times x(t-1) + \epsilon(t). \tag{1}$$

γ_1 is interpreted as a measure of how significant x is in predicting the equity premium. The most prominent x variables explored in the literature are the dividend price ratio and dividend yield, the earnings price ratio and dividend-earnings (payout) ratio, various interest rates and spreads, the inflation rates, the book-to-market ratio, volatility, the investment-capital ratio, the consumption, wealth, and income ratio, and aggregate net or equity issuing activity.

The literature is difficult to absorb. Different articles use different techniques, variables, and time periods. Results from articles that were written years ago may change when more recent data is used. Some articles

Thanks to Malcolm Baker, Ray Ball, John Campbell, John Cochrane, Francis Diebold, Ravi Jagannathan, Owen Lamont, Sydney Ludvigson, Rajnish Mehra, Michael Roberts, Jay Shanken, Samuel Thompson, Jeff Wurgler, and Yihong Xia for comments, and Todd Clark for providing us with some critical McCracken values. We especially appreciate John Campbell and Sam Thompson for challenging our earlier drafts, and iterating mutually over working papers with opposite perspectives. Address correspondence to Amit Goyal, http://www.goizueta.emory.edu/agoyal E-mail: mailto:amit_goyal@bus.emory.edu http://welch.econ.brown.edu E-mail: mailto:ivo_welch@brown.edu, or e-mail: amit_goyal@bus.emory.edu.

doi:10.1093/rfs/hhm014 Advance Access publication March 17, 2007

contradict the findings of others. Still, most readers are left with the impression that "prediction works"—though it is unclear exactly what works. The prevailing tone in the literature is perhaps best summarized by Lettau and Ludvigson (2001, p.842)

> "It is now widely accepted that excess returns are predictable by variables such as dividend-price ratios, earnings-price ratios, dividend-earnings ratios, and an assortment of other financial indicators."

There are also a healthy number of current article that further cement this perspective and a large theoretical and normative literature has developed that stipulates how investors should allocate their wealth as a function of the aforementioned variables.

The goal of our own article is to comprehensively re-examine the empirical evidence as of early 2006, evaluating each variable using the same methods (mostly, but not only, in linear models), time-periods, and estimation frequencies. The evidence suggests that most models are unstable or even spurious. Most models are no longer significant even in-sample (IS), and the few models that still are usually fail simple regression diagnostics. Most models have performed poorly for over 30 years IS. For many models, any earlier apparent statistical significance was often based exclusively on years up to *and especially on* the years of the Oil Shock of 1973–1975. Most models have poor out-of-sample (OOS) performance, but not in a way that merely suggests lower power than IS tests. They predict poorly late in the sample, not early in the sample. (For many variables, we have difficulty finding robust statistical significance even when they are examined only during their most favorable contiguous OOS sub-period.) Finally, the OOS performance is not only a useful model diagnostic for the IS regressions but also interesting in itself for an investor who had sought to use these models for market-timing. Our evidence suggests that the models would not have helped such an investor.

Therefore, although it is possible to search for, to occasionally stumble upon, and then to defend some seemingly statistically significant models, we interpret our results to suggest that a healthy skepticism is appropriate when it comes to predicting the equity premium, at least as of early 2006. The models do not seem robust.

Our article now proceeds as follows. We describe our data—available at the RFS website—in Section 1 and our tests in Section 2. Section 3 explores our base case—predicting equity premia annually using OLS forecasts. In Sections 4 and 5, we predict equity premia on 5-year and monthly horizons, the latter with special emphasis on the suggestions in Campbell and Thompson (2005). Section 6 tries earnings and dividend ratios with longer memory as independent variables, corrections for persistence in

regressors, and encompassing model forecasts. Section 7 reviews earlier literature. Section 8 concludes.

1. Data Sources and Data Construction

Our dependent variable is always the equity premium, that is, the total rate of return on the stock market minus the prevailing short-term interest rate.

Stock Returns : We use S&P 500 index returns from 1926 to 2005 from Center for Research in Security Press (CRSP) month-end values. Stock returns are the continuously compounded returns on the S&P 500 index, including dividends. For yearly and longer data frequencies, we can go back as far as 1871, using data from Robert Shiller's website. For monthly frequency, we can only begin in the CRSP period, that is, 1927.

Risk-free Rate : The risk-free rate from 1920 to 2005 is the Treasury-bill rate. Because there was no risk-free short-term debt prior to the 1920s, we had to estimate it. Commercial paper rates for New York City are from the National Bureau of Economic Research (NBER) Macrohistory data base. These are available from 1871 to 1970. We estimated a regression from 1920 to 1971, which yielded

$$\text{Treasury-bill rate} = -0.004 + 0.886 \times \text{Commercial Paper Rate}, \quad (2)$$

with an R^2 of 95.7%. Therefore, we instrumented the risk-free rate from 1871 to 1919 with the predicted regression equation. The correlation for the period 1920 to 1971 between the equity premium computed using the actual Treasury-bill rate and that computed using the predicted Treasury-bill rate (using the commercial paper rate) is 99.8%.

The equity premium had a mean (standard deviation) of 4.85% (17.79%) over the entire sample from 1872 to 2005; 6.04% (19.17%) from 1927 to 2005; and 4.03% (15.70%) from 1965 to 2005.

Our first set of independent variables are primarily stock characteristics:

Dividends : Dividends are 12-month moving sums of dividends paid on the S&P 500 index. The data are from Robert Shiller's website from 1871 to 1987. Dividends from 1988 to 2005 are from the S&P Corporation. The *Dividend Price Ratio* (**d/p**) is the difference between the log of dividends and the log of prices. The Dividend Yield (**d/y**) is the difference between the log of dividends and the log of *lagged* prices. [See, e.g., Ball (1978), Campbell (1987), Campbell and Shiller (1988a, 1988b), Campbell and Viceira (2002), Campbell and Yogo (2006), the survey in Cochrane (1997), Fama and French (1988), Hodrick (1992), Lewellen (2004), Menzly, Santos, and Veronesi (2004), Rozeff (1984), and Shiller (1984).]

Earnings : Earnings are 12-month moving sums of earnings on the S&P 500 index. The data are again from Robert Shiller's website from 1871 to 1987. Earnings from 1988 to 2005 are our own estimates based on

interpolation of quarterly earnings provided by the S&P Corporation. The *Earnings Price Ratio* (**e/p**) is the difference between the log of earnings and the log of prices. (We also consider variations, in which we explore multiyear moving averages of numerator or denominator, e.g., as in $\mathbf{e^{10}/p}$, which is the moving ten-year average of earnings divided by price.) The *Dividend Payout Ratio* (**d/e**) is the difference between the log of dividends and the log of earnings. [See, e.g., Campbell and Shiller (1988a, 1998) and Lamont (1998).]

Stock Variance (svar) : Stock Variance is computed as sum of squared daily returns on the S&P 500. G. William Schwert provided daily returns from 1871 to 1926; data from 1926 to 2005 are from CRSP. [See Guo (2006).]

Cross-Sectional Premium (csp) : The cross-sectional beta premium measures the relative valuations of high- and low-beta stocks and is proposed in Polk, Thompson, and Vuolteenaho (2006). The **csp** data are from Samuel Thompson from May 1937 to December 2002.

Book Value : Book values from 1920 to 2005 are from Value Line's website, specifically their *Long-Term Perspective Chart* of the Dow Jones Industrial Average. The *Book-to-Market Ratio* (**b/m**) is the ratio of book value to market value for the Dow Jones Industrial Average. For the months from March to December, this is computed by dividing book value at the end of the previous year by the price at the end of the current month. For the months of January and February, this is computed by dividing book value at the end of two years ago by the price at the end of the current month. [See, e.g, Kothari and Shanken (1997) and Pontiff and Schall (1998).]

Corporate Issuing Activity : We entertain two measures osf corporate issuing activity. *Net Equity Expansion* (**ntis**) is the ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks. This dollar amount of net equity issuing activity (IPOs, SEOs, stock repurchases, less dividends) for NYSE listed stocks is computed from CRSP data as

$$\text{Net Issue}_t \;=\; \text{Mcap}_t - \text{Mcap}_{t-1} \times (1 + \text{vwretx}_t), \tag{3}$$

where Mcap is the total market capitalization, and vwretx is the value weighted return (excluding dividends) on the NYSE index.[1] These data are available from 1926 to 2005. **ntis** is closely related, but not identical, to a variable proposed in Boudoukh, Michaely, Richardson, and Roberts (2007). The second measure, *Percent Equity Issuing* (**eqis**), is the ratio of equity issuing activity as a fraction of total issuing activity. This is the variable proposed in Baker and Wurgler (2000). The authors provided us with the data, except for 2005, which we added ourselves. The first equity

[1] This calculation implicitly assumes that the delisting return is −100 percent. Using the actual delisting return, where available, or ignoring delistings altogether, has no impact on our results.

issuing measure is relative to aggregate market cap, while the second is relative to aggregate corporate issuing.

Our next set of independent variables is interest-rate related:

Treasury Bills (tbl) : Treasury-bill rates from 1920 to 1933 are the *U.S. Yields On Short-Term United States Securities, Three-Six Month Treasury Notes and Certificates, Three Month Treasury* series in the NBER Macrohistory data base. Treasury-bill rates from 1934 to 2005 are the *3-Month Treasury Bill: Secondary Market Rate* from the economic research data base at the Federal Reserve Bank at St. Louis (FRED. [See, e.g., Campbell (1987) and Hodrick (1992).]

Long Term Yield (lty) : Our long-term government bond yield data from 1919 to 1925 is the *U.S. Yield On Long-Term United States Bonds* series in the NBER's Macrohistory data base. Yields from 1926 to 2005 are from Ibbotson's *Stocks, Bonds, Bills and Inflation Yearbook*, the same source that provided the *Long Term Rate of Returns* (**ltr**). The Term Spread (**tms**) is the difference between the long term yield on government bonds and the Treasury-bill. [See, e.g., Campbell (1987) and Fama and French (1989).]

Corporate Bond Returns : Long-term corporate bond returns from 1926 to 2005 are again from Ibbotson's *Stocks, Bonds, Bills and Inflation Yearbook*. *Corporate Bond Yields* on AAA and BAA-rated bonds from 1919 to 2005 are from FRED. The *Default Yield Spread* (**dfy**) is the difference between BAA and AAA-rated corporate bond *yields*. The *Default Return Spread* (**dfr**) is the difference between long-term corporate bond and long-term government bond *returns*. [See, e.g., Fama and French (1989) and Keim and Stambaugh (1986).]

Inflation (infl) : Inflation is the *Consumer Price Index (All Urban Consumers)* from 1919 to 2005 from the Bureau of Labor Statistics. Because inflation information is released only in the following month, we wait for one month before using it in our monthly regressions. [See, e.g., Campbell and Vuolteenaho (2004), Fama (1981), Fama and Schwert (1977), and Lintner (1975).]

Like inflation, our next variable is also a common broad macroeconomic indicator.

Investment to Capital Ratio (i/k) : The investment to capital ratio is the ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the whole economy. This is the variable proposed in Cochrane (1991). John Cochrane kindly provided us with updated data.

Of course, many articles explore multiple variables. For example, Ang and Bekaert (2003) explore both interest rate and dividend related variables. In addition to simple univariate prediction models, we also entertain two methods that rely on multiple variables (**all** and **ms**), and two models that are rolling in their independent variable construction (**cay** and **ms**).

A "Kitchen Sink" Regression (all): This includes all the aforementioned variables. (It does not include **cay**, described below, partly due to limited data availability of **cay**.)
Consumption, wealth, income ratio (cay): Lettau and Ludvigson (2001) estimate the following equation:

$$c_t = \alpha + \beta_a \cdot a_t + \beta_y \cdot y_t + \sum_{i=-k}^{k} b_{a,i} \cdot \Delta a_{t-i} + \sum_{i=-k}^{k} b_{y,i} \cdot \Delta y_{t-i} + \epsilon_t, \quad t = k+1, \ldots, T-k, \tag{4}$$

where c is the aggregate consumption, a is the aggregate wealth, and y is the aggregate income. Using estimated coefficients from the above equation provides $\mathbf{cay} \equiv \widehat{\text{cay}}_t = c_t - \hat{\beta}_a \cdot a_t - \hat{\beta}_y \cdot y_t, \quad t = 1, \ldots, T$. Note that, unlike the estimation equation, the fitting equation does not use look-ahead data. Eight leads/lags are used in quarterly estimation ($k = 8$) while two lags are used in annual estimation ($k = 2$). [For further details, see Lettau and Ludvigson (2001).] Data for **cay**'s construction are available from Martin Lettau's website at quarterly frequency from the second quarter of 1952 to the fourth quarter of 2005. Although annual data from 1948 to 2001 is also available from Martin Lettau's website, we reconstruct the data following their procedure as this allows us to expand the time-series from 1945 to 2005 (an addition of 7 observations).

Because the Lettau–Ludvigson measure of **cay** is constructed using look-ahead (in-sample) estimation regression coefficients, we also created an equivalent measure that excludes advance knowledge from the estimation equation and thus uses only prevailing data. In other words, if the current time period is 's', then we estimated Equation (4) using only the data up to 's' through

$$c_t = \alpha + \beta_a^s \cdot a_t + \beta_y^s \cdot y_t + \sum_{i=-k}^{k} b_{a,i}^s \cdot \Delta a_{t-i} + \sum_{i=-k}^{k} b_{y,i}^s \cdot \Delta y_{t-i} + \epsilon_t, \; t = k+1, \ldots, s-k, \tag{5}$$

This measure is called **caya** ("ante") to distinguish it from the traditional variable **cayp** constructed with look-ahead bias ("post"). The superscript on the betas indicates that these are rolling estimates, that is, a set of coefficients used in the construction of one $\mathbf{caya}_S$ measure in one period.
A **model selection** approach, named "**ms**." If there are K variables, we consider 2^K models essentially consisting of all possible combinations

of variables. (As with the kitchen sink model, **cay** is not a part of the **ms** selection.) Every period, we select one of these models that gives the minimum cumulative prediction errors up to time t. This method is based on Rissanen (1986) and is recommended by Bossaerts and Hillion (1999). Essentially, this method uses our criterion of minimum OOS prediction errors to choose among competing models *in each time period* t. This is also similar in spirit to the use of a more conventional criterion (like R^2) in Pesaran and Timmermann (1995) (who do not entertain our NULL hypothesis). This selection model also shares a certain flavor with our encompassing tests in Section 6, where we seek to find an optimal rolling combination between each model and an unconditional historical equity premium average, and with the Bayesian model selection approach in Avramov (2002).

The latter two models, **cay** and **ms**, are revised every period, which render IS regressions problematic. This is also why we did not include **caya** in the kitchen sink specification.

2. Empirical Procedure

Our base regression coefficients are estimated using OLS, although statistical significance is always computed from bootstrapped F-statistics (taking correlation of independent variables into account).

OOS statistics: The OOS forecast uses only the data available up to the time at which the forecast is made. Let e_N denote the vector of rolling OOS errors from the historical mean model and e_A denote the vector of rolling OOS errors from the OLS model. Our OOS statistics are computed as

$$R^2 = 1 - \frac{\text{MSE}_A}{\text{MSE}_N}, \quad \overline{R}^2 = R^2 - (1 - R^2) \times \left(\frac{T-k}{T-1}\right),$$
$$\Delta\text{RMSE} = \sqrt{\text{MSE}_N} - \sqrt{\text{MSE}_A},$$
$$\text{MSE-F} = (T - h + 1) \times \left(\frac{\text{MSE}_N - \text{MSE}_A}{\text{MSE}_A}\right), \tag{6}$$

where h is the degree of overlap ($h = 1$ for no overlap). MSE-F is McCracken's (2004) F-statistic. It tests for equal MSE of the unconditional forecast and the conditional forecast (i.e., $\Delta\text{MSE} = 0$).[2] We generally do

[2] Our earlier drafts also entertained another performance metric, the mean absolute error difference ΔMAE. The results were similar. These drafts also described another OOS-statistic, MSE-T $= \sqrt{T+1-2\cdot h+h\cdot(h-1)/T}\cdot\left[\frac{\overline{d}}{\widehat{se}(\overline{d})}\right]$, where $d_t = e_{Nt} - e_{At}$, and $\overline{d} = T^{-1}\cdot\sum_t^T d_t = \text{MSE}_N - \text{MSE}_A$ over the entire OOS period, and T is the total number of forecast observations. This is the Diebold and Mariano (1995) t-statistic modified by Harve, Leybourne, and Newbold (1997). (We still use the latter as bounds in our plots, because we know the full distribution.) Again, the results were similar. We chose to use the MSE-F in this article because Clark and McCracken (2001) find that MSE-F has higher power than MSE-T.

not report MSE-F statistics, but instead use their bootstrapped critical levels to provide statistical significance levels via stars in the tables.

For our encompassing tests in Section 6, we compute

$$\text{ENC} = \frac{T-h+1}{T} \times \frac{\sum_{t=1}^{T}\left(e_{Nt}^2 - e_{Nt}\cdot e_{At}\right)}{\text{MSE}_A}, \tag{7}$$

which is proposed by Clark and McCracken (2001). They also show that the MSE-F and ENC statistics follow nonstandard distributions when testing nested models, because the asymptotic difference in squared forecast errors is exactly 0 with 0 variance under the NULL, rendering the standard distributions asymptotically invalid. Because our models are nested, we could use asymptotic critical values for MSE tests provided by McCracken, and asymptotic critical values for ENC tests provided by Clark and McCracken. However, because we use relatively small samples, because our independent variables are often highly serially correlated, and especially because we need critical values for our 5-year *overlapping* observations (for which asymptotic critical values are not available), we obtain critical values from the bootstrap procedure described below. (The exceptions are that critical values for **caya**, **cayp**, and **all** models are not calculated using a bootstrap, and critical values for **ms** model are not calculated at all.) The NULL hypothesis is that the unconditional forecast is not inferior to the conditional forecast, so our critical values for OOS test are for a one-sided test (critical values of IS tests are, as usual, based on two-sided tests).[3]

Bootstrap : Our bootstrap follows Mark (1995) and Kilian (1999) and imposes the NULL of no predictability for calculating the critical values. In other words, the data generating process is assumed to be

$$\begin{aligned} y_{t+1} &= \alpha \qquad\qquad + u_{1t+1} \\ x_{t+1} &= \mu + \rho \times x_t + u_{2t+1}. \end{aligned}$$

The bootstrap for calculating power assumes the data generating process is

$$\begin{aligned} y_{t+1} &= \alpha + \beta \times x_t + u_{1t+1} \\ x_{t+1} &= \mu + \rho \times x_t + u_{2t+1}, \end{aligned}$$

where both β and ρ are estimated by OLS using the full sample of observations, with the residuals stored for sampling. We then generate

[3] If the regression coefficient β is small (so that explanatory power is low or the IS R^2 is low), it may happen that our unconditional model outperforms on OOS because of estimation error in the rolling estimates of β. In this case, ΔRMSE might be negative but still significant *because these tests are ultimately tests of whether β is equal to zero*.

10,000 bootstrapped time series by drawing with replacement from the residuals. The initial observation—preceding the sample of data used to estimate the models—is selected by picking one date from the actual data at random. This bootstrap procedure not only preserves the autocorrelation structure of the predictor variable, thereby being valid under the Stambaugh (1999) specification, but also preserves the cross-correlation structure of the two residuals.[4]

Statistical Power: Our article entertains both IS and OOS tests. Inoue and Kilian (2004) show that the OOS tests used in this paper are less powerful than IS tests, even though their size properties are roughly the same. Similar critiques of the OOS tests in our article have been noted by Cochrane (2005) and Campbell and Thompson (2005). We believe this is the wrong way to look at the issue of power for two reasons:

(i) It is true that under a well-specified, stable underlying model, an IS OLS estimator is more efficient. Therefore, a researcher who has complete confidence in her underlying model specification (but not the underlying model parameters) should indeed rely on IS tests to establish significance—the alternative to OOS tests does have lower power. However, the point of any regression diagnostics, such as those for heteroskedasticity and autocorrelation, is always to subject otherwise seemingly successful regression models to a number of reasonable diagnostics when there is some model uncertainty. Relative to not running the diagnostic, by definition, any diagnostic that can reject the model at this stage sacrifices power *if* the specified underlying model is correct. In our forecasting regression context, OOS performance just happens to be one natural and especially useful diagnostic statistic. It can help determine whether a model is stable and wellspecified, or changing over time, either suddenly or gradually.

This also suggests why the simple power experiment performed in some of the aforementioned critiques of our own paper is wrong. It is unreasonable to propose a model if the IS performance is insignificant, regardless of its OOS performance. Reasonable (though not necessarily statistically significant) OOS performance is not a substitute, but a necessary complement for IS performance in order to establish the quality of the underlying model specification. The thought experiments and analyses in the critiques, which simply compare the power of OOS tests to that of IS tests, especially under their assumption of a correctly specified stable model, is therefore incorrect. The correct power

[4] We do not bootstrap for **cayp** because it is calculated using *ex-post* data; for **caya** and **ms** because these variables change each period; and for **all** because of computational burden.

experiment instead should explore whether *conditional on observed IS significance*, OOS diagnostics are reasonably powerful. We later show that they are.

Not reported in the tables, we also used the CUSUMQ test to test for model stability. Although this is a weak test, we can reject stability for all monthly models: and for all annual models except for **ntis**, **i/k**, and **cayp**, when we use data beginning in 1927. Thus, the CUSUMQ test sends the same message about the models as the findings that we shall report.

(ii) All of the OOS tests in our paper do not fail in the way the critics suggest. Low-power OOS tests would produce relatively poor predictions early and relatively good predictions late in the sample. Instead, all of our models show the opposite behavior—good OOS performance early, bad OOS performance late.

A simple alternative OOS estimator, which downweights early OOS predictions relative to late OOS predictions, would have more power than our unweighted OOS prediction test. Such a modified estimator would both be more powerful, *and* it would show that all models explored in our article perform even worse. (We do not use it only to keep it simple and to avoid a "cherry-picking-the-test" critique.)

Estimation Period : It is not clear how to choose the periods over which a regression model is estimated and subsequently evaluated. This is even more important for OOS tests. Although any choice is necessarily ad-hoc in the end, the criteria are clear. It is important to have enough initial data to get a reliable regression estimate at the start of evaluation period, and it is important to have an evaluation period that is long enough to be representative. We explore three time period specifications: the first begins OOS forecasts 20 years after data are available; the second begins OOS forecast in 1965 (or 20 years after data are available, whichever comes later); the third ignores all data prior to 1927 even in the estimation.[5] If a variable does not have complete data, some of these time-specifications can overlap. Using three different periods reflects different trade-offs between the desire to obtain statistical power and the desire to obtain results that remain relevant today. In our graphical analysis later, we also evaluate the rolling predictive performance of variables. This analysis helps us identify periods of superior or inferior performance and can be seen as invariant to the choice of the OOS evaluation period (though not to the choice of the estimation period).

[5] We also tried estimating our models only with data after World War II, as recommended by Lewellen (2004). Some properties in some models change, especially when it comes to statistical significance and the importance of the Oil Shock for one variable, **d/p**. However, the overall conclusions of our article remain.

3. Annual Prediction

Table 1 shows the predictive performance of the forecasting models on annual forecasting horizons. Figures 1 and 2 graph the IS and OOS performance of variables in Table 1. For the IS regressions, the performance is the cumulative squared demeaned equity premium minus the cumulative squared regression residual. For the OOS regressions, this is the cumulative squared prediction errors of the prevailing mean minus the cumulative squared prediction error of the predictive variable from the linear historical regression. Whenever a line increases, the ALTERNATIVE predicted better; whenever it decreases, the NULL predicted better. The units in the graphs are not intuitive, but the time-series pattern allows diagnosis of years with good or bad performance. Indeed, the final ΔSSE statistic in the OOS plot is sign-identical with the ΔRMSE statistic in our tables. The standard error of all the observations in the graphs is based on translating MSE-T statistic into symmetric 95% confidence intervals based on the McCracken (2004) critical values; the tables differ in using the MSE-F statistic instead.

The reader can easily adjust perspective to see how variations in starting or ending date would impact the conclusion—by shifting the graph up or down (redrawing the $y = 0$ horizontal zero line). Indeed, a horizontal line and the right-side scale indicate the equivalent zero-point for the second time period specification, in which we begin forecasts in 1965 (this is marked "Start=1965 Zero Val" line). The plots have also vertically shifted the IS errors, so that the IS line begins at zero on the date of our first OOS prediction. The Oil Shock recession of 1973 to 1975, as identified by the NBER, is marked by a vertical (red) bar in the figures.[6]

In addition to the figures and tables, we also summarize models' performances in small in-text summary tables, which give the IS-$\overline{R}^2$ and OOS-$\overline{R}^2$ for two time periods: the most recent 30 years and the entire sample period. The $\overline{R}^2$ for the subperiod is not the $\overline{R}^2$ for a different model estimated only over the most recent three decades, but the residual fit for the overall model over the subset of data points (e.g., computed simply as 1-SSE/SST for the last 30 years' residuals). The most recent three decades after the Oil Shock can help shed light on whether a model is likely to still perform well nowadays. Generally, it is easiest to understand the data by looking first at the figures, then at the in-text table, and finally at the full table.

A well-specified signal would inspire confidence in a potential investor if it had

[6] The actual recession period was from November 1973 to March 1975. We treat both 1973 and 1975 as years of Oil Shock recession in annual prediction.

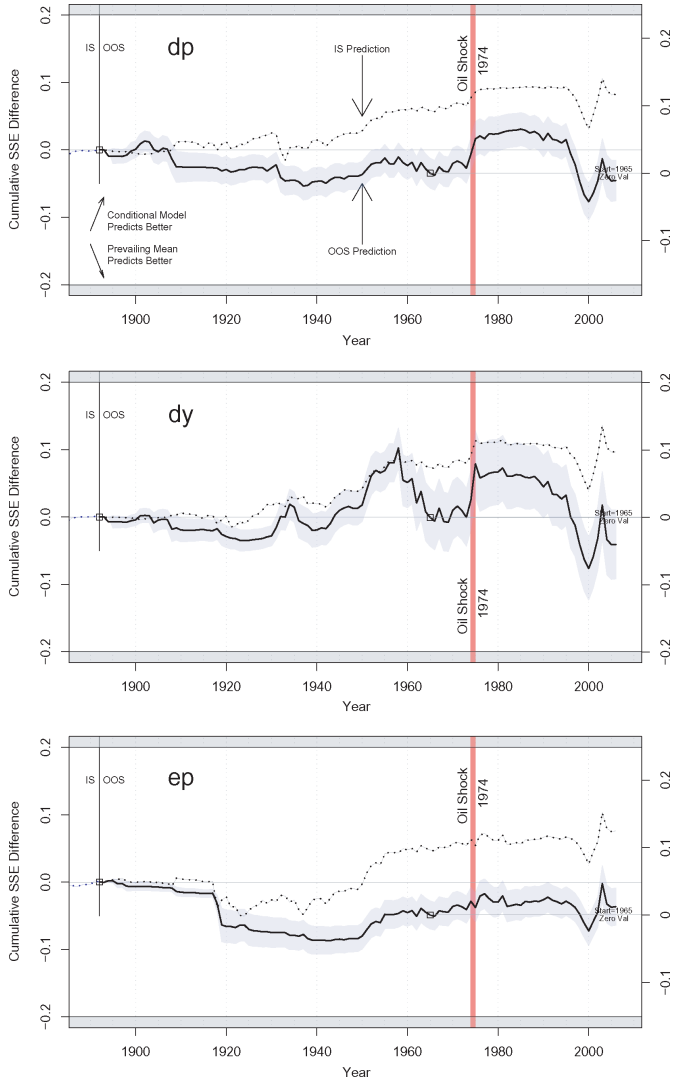

Figure 1
Annual performance of IS insignificant predictors.
Explanation: These figures plot the IS and OOS performance of annual predictive regressions. Specifically, these are the cumulative squared prediction errors of the NULL minus the cumulative squared prediction error of the ALTERNATIVE. The ALTERNATIVE is a model that relies on predictive variables noted in each graph. The NULL is the prevailing equity premium mean for the OOS graph, and the full-period equity premium mean for the IS graph. The IS prediction relative performance is dotted (and usually above), the OOS prediction relative perfomance is solid. An increase in a line indicates better performance of the named model; a decrease in a line indicates better performance of the NULL. The blue band is the equivalent of 95% two-sided levels, based on MSE-T critical values from McCracken (2004). (MSE-T is the Diebold and Mariano (1995) *t*-statistic modified by Harvey, Leybourne, and Newbold (1998)). The right axis shifts the zero point to 1965. The Oil Shock is marked by a red vertical line.

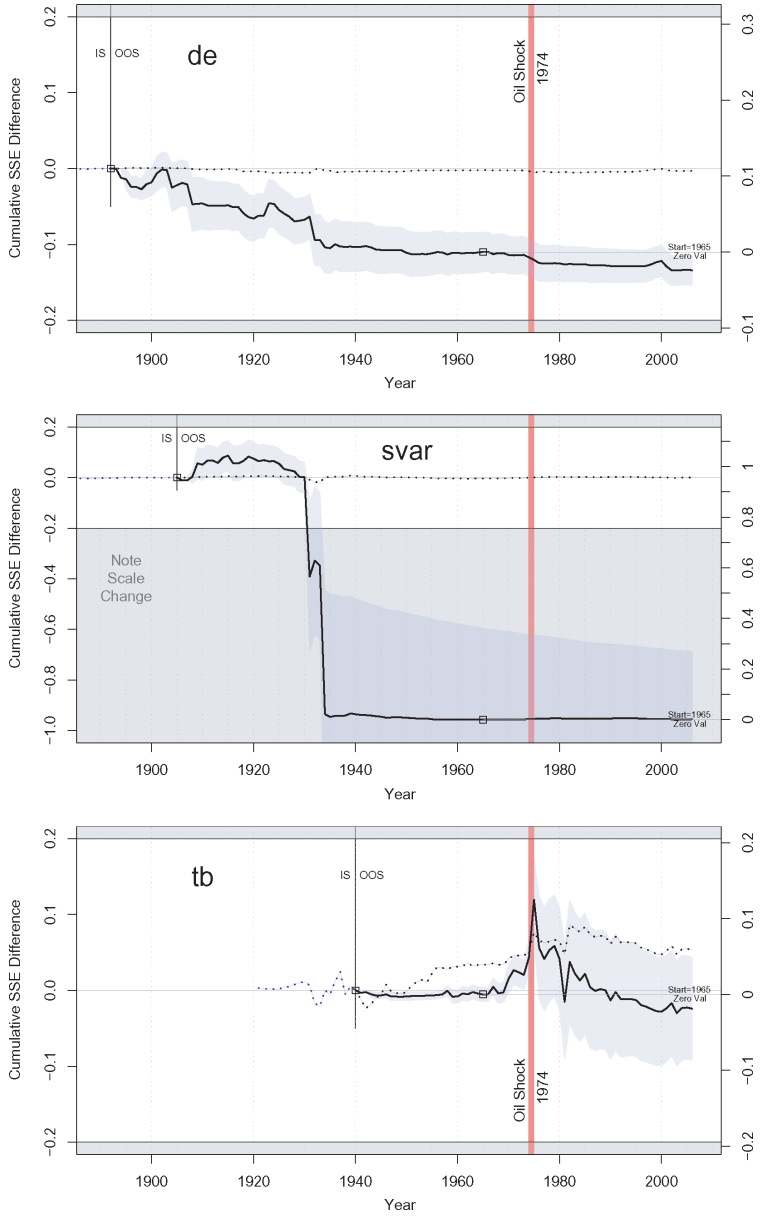

Figure 1 Continued

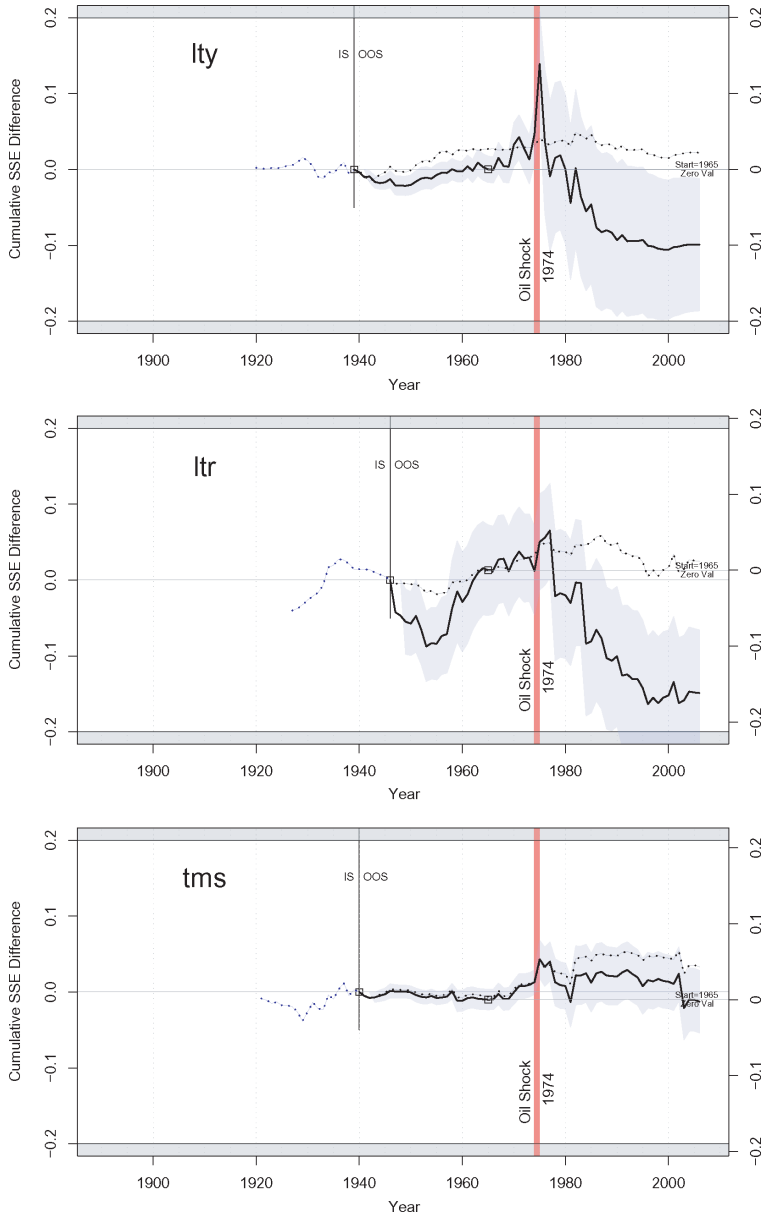

Figure 1 Continued

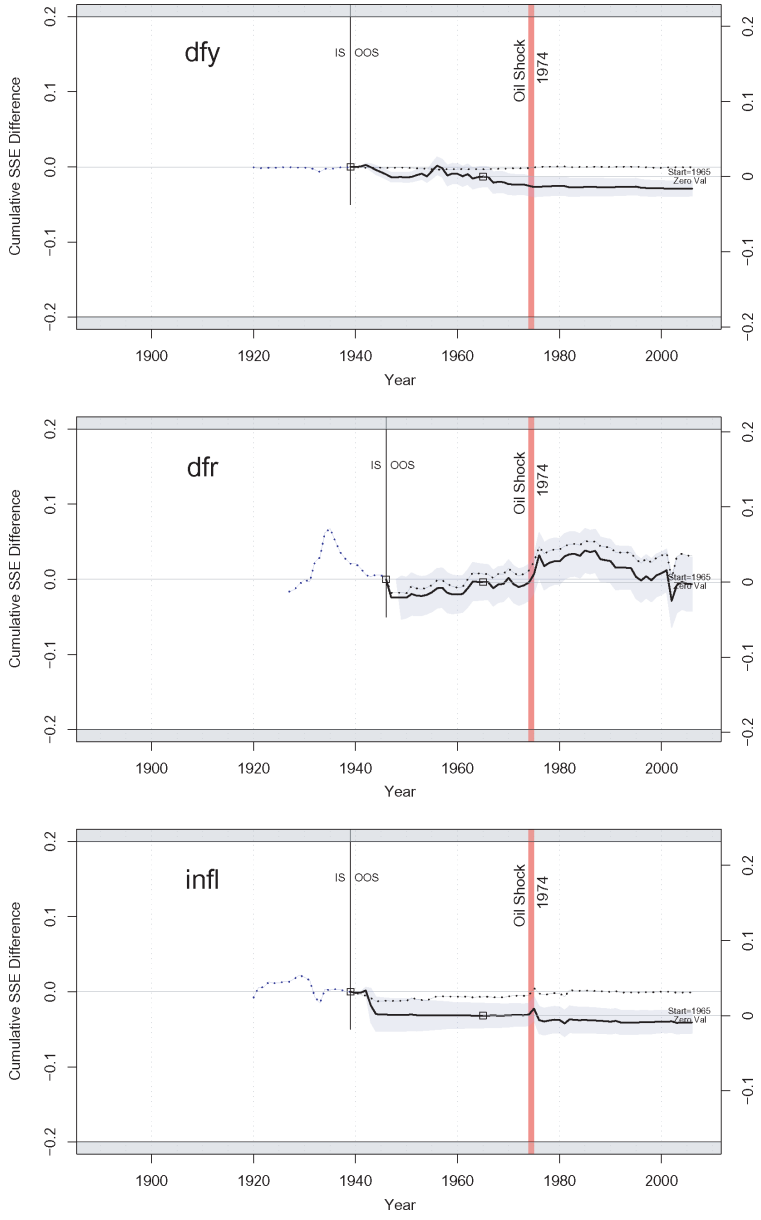

Figure 1 Continued

Table 1
Forecasts at annual frequency

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for log equity premium forecasts at annual frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. Stock returns are price changes, including dividends, of the S&P500. All numbers are in percent per year, except except $\overline{R}^2$ and power which are simple percentages. A star next to IS-$\overline{R}^2$ denotes significance of the in-sample regression as measured by F-statistics (critical values of which are obtained empirically from bootstrapped distributions). The column 'IS for OOS' gives the IS-$\overline{R}^2$ for the OOS period. ΔRMSE is the RMSE (root mean square error) difference between the unconditional forecast and the conditional forecast for the same sample/forecast period. Positive numbers signify superior out-of-sample conditional forecast. The OOS-$\overline{R}^2$ is defined in Equation 6. A star next to OOS-$\overline{R}^2$ is based on significance of MSE-F statistic by McCracken (2004), which tests for equal MSE of the unconditional forecast and the conditional forecast. One-sided critical values of MSE statistics are obtained empirically from bootstrapped distributions, except for caya and all models where they are obtained from McCracken (2004). Critical values for the ms model are not calculated. Power is calculated as the fraction of draws where the simulated ΔRMSE is greater than the empirically calculated 95% critical value. The two numbers under the power column are for all simulations and for those simulations in which the in-sample estimate was significant at the 95% level. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

			Full Sample									1927–2005
				Forecasts begin 20 years after sample				Forecasts begin 1965				Sample
			IS	IS for	OOS			IS for	OOS			IS
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	$\overline{R}^2$
Full Sample, Not Significant IS												
dfy	Default yield spread	1919–2005	−1.18		−3.29	−0.14			−4.15	−0.12		−1.31
infl	Inflation	1919–2005	−1.00		−4.07	−0.20			−3.56	−0.08		−0.99
svar	Stock variance	1885–2005	−0.76		−27.14	−2.33			−2.44	+0.01		−1.32
d/e	Dividend payout ratio	1872–2005	−0.75		−4.33	−0.31			−4.99	−0.18		−1.24
lty	Long term yield	1919–2005	−0.63		−7.72	−0.47			−12.57	−0.76		−0.94
tms	Term spread	1920–2005	0.16		−2.42	−0.07			−2.96	−0.03		0.89
tbl	Treasury-bill rate	1920–2005	0.34		−3.37	−0.14			−4.90	−0.18		0.15
dfr	Default return spread	1926–2005	0.40		−2.16	−0.03			−2.82	−0.02		0.32
d/p	Dividend price ratio	1872–2005	0.49		−2.06	−0.11			−3.69	−0.09		1.67
d/y	Dividend yield	1872–2005	0.91		−1.93	−0.10			−6.68	−0.31		**2.71***
ltr	Long term return	1926–2005	0.99		−11.79	−0.76			−18.38	−1.18		0.92
e/p	Earning price ratio	1872–2005	1.08		−1.78	−0.08			−1.10	0.11		**3.20***

Table 1
Continued

			Full Sample									1927–2005
				Forecasts begin 20 years after sample				Forecasts begin 1965				Sample
			IS	IS for	OOS			IS for	OOS			IS
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	$\overline{R}^2$
Full Sample, Significant IS												
b/m	Book to market	1921–2005	**3.20***	1.13	−1.72	−0.01	42 (67)	−7.29	−12.71	−0.77	40 (61)	**4.14***
i/k	Invstmnt capital ratio	1947–2005	**6.63****	−0.25	−1.77	0.07	47 (77)		Same			Same
ntis	Net equity expansion	1927–2005	**8.15*****	−4.21	−5.07	−0.26	57 (78)	0.96	−6.79	−0.32	53 (72)	Same
eqis	Pct equity issuing	1927–2005	**9.15*****	2.81	**2.04****	0.30	72 (85)	3.64	−1.00	0.12	66 (77)	Same
all	Kitchen sink	1927–2005	**13.81****	2.62	−139.03	−5.97	- (-)	−20.91	−176.18	−6.19	- (-)	Same
Full sample, no IS equivalent (caya, ms) or Ex-Post Information (cayp)												
cayp	Cnsmptn, wlth, incme	1945–2005	**15.72*****	20.70	**16.78*****	1.61	- (-)		Same			Same
caya	Cnsmptn, wlth, incme	1945–2005	-	-	−4.33	−0.14	- (-)		Same			Same
ms	Model selection	1927–2005	-	-	−22.50	−1.69	- (-)	-	−23.71	−1.79	- (-)	Same
1927-2005 Sample, Significant IS												Full Sample
d/y	Dividend yield	1927–2005	**2.71***					−0.35	−6.44	−0.30	30 (71)	0.91
e/p	Earning price ratio	1927–2005	**3.20***					−0.94	−3.15	−0.05	39 (64)	1.08
b/m	Book to market	1927–2005	**4.14***					−8.65	−19.46	−1.26	45 (64)	3.20*

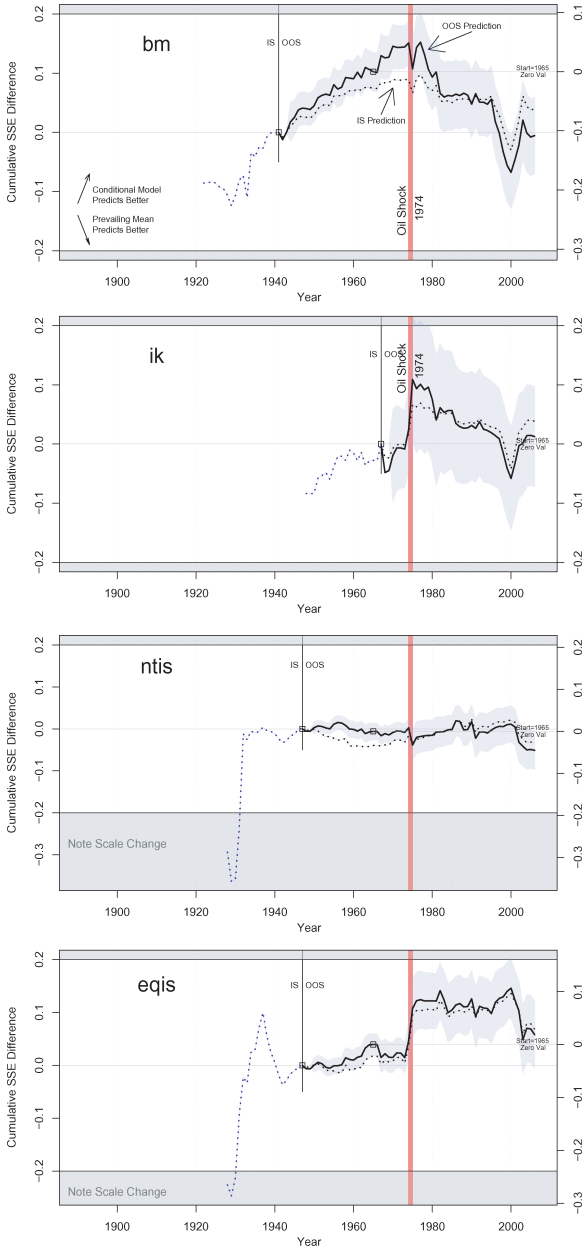

Figure 2
Annual performance of predictors that are not in-sample significant
Explanation: See Figure 1.

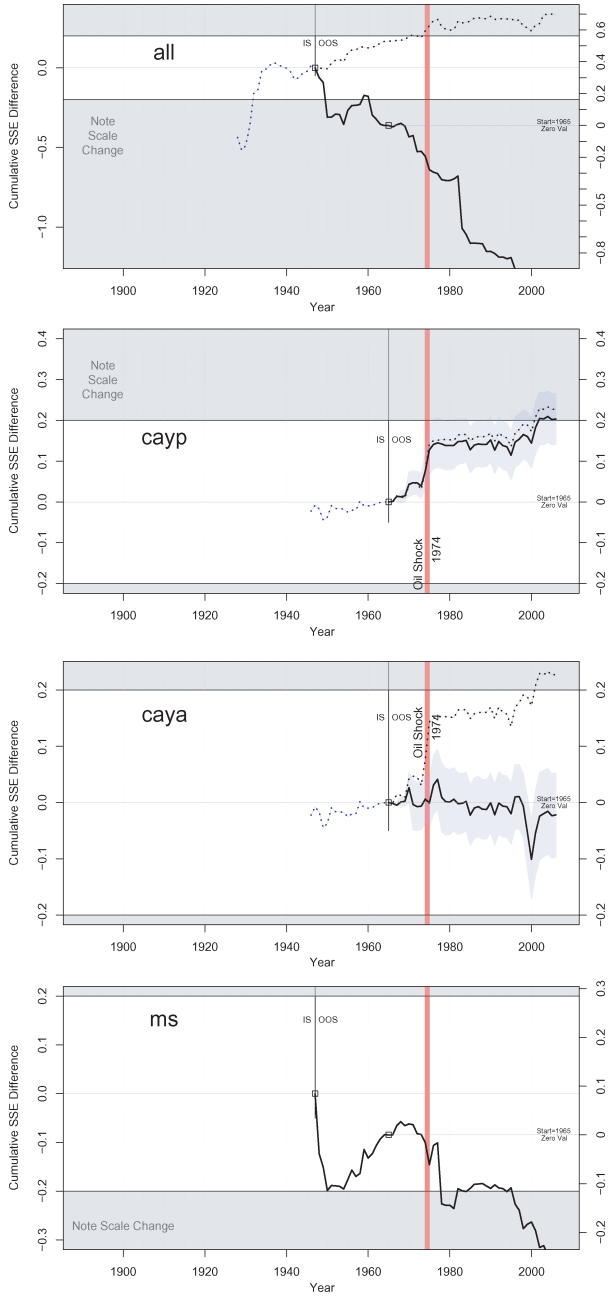

Figure 2 Continued

(i) both significant IS and reasonably good OOS performance over the entire sample period;
(ii) a generally upward drift (of course, an irregular one);
(iii) an upward drift which occurs not just in one short or unusual sample period—say just the two years around the Oil Shock;
(iv) an upward drift that remains positive over the most recent several decades—otherwise, even a reader taking the long view would have to be concerned with the possibility that the underlying model has drifted.

There are also other diagnostics that stable models should pass (heteroskedasticity, residual autocorrelation, etc.), but we do not explore them in our article.

3.1 In-sample insignificant models

As already mentioned, if a model has no IS performance, its OOS performance is not interesting. However, because some of the IS insignificant models are so prominent, and because it helps to understand why they may have been considered successful forecasters in past articles, we still provide some basic statistics and graph their OOS performance. The most prominent such models are the following:

Dividend Price Ratio: Figure 1 shows that there were four distinct periods for the **d/p** model, and this applies both to IS and OOS performance. **d/p** had mild underperformance from 1905 to WW II, good performance from WW II to 1975, neither good nor bad performance until the mid-1990s, and poor performance thereafter. The best sample period for **d/p** was from the mid 1930s to the mid-1980s. For the OOS, it was 1937 to 1984, although over half of the OOS performance was due to the Oil Shock. Moreover, the plot shows that the OOS performance of the **d/p** regression was consistently worse than the performance of its IS counterpart. The distance between the IS and OOS performance increased steadily until the Oil Shock.

Over the most recent 30 years (1976 to 2005), **d/p**'s performance is negative both IS and OOS. Over the entire period, **d/p** underperformed the prevailing mean OOS, too:

d/p	Recent 30 years	All years
IS $\overline{R}^2$	−4.80%	0.49%
OOS $\overline{R}^2$	−15.14%	−2.06%

Dividend Yield : Figure 1 shows that the **d/y** model's IS patterns look broadly like those of **d/p**. However, its OOS pattern was much more volatile: **d/y** predicted equity premia well during the Great Depression (1930 to 1933), the period from World War II to 1958, the Oil Shock of

1973–1975, and the market decline of 2000–2002. It had large prediction errors from 1958 to 1965 and from 1995 to 2000, and it had unremarkable performance in other years. The best OOS sample period started around 1925 and ended either in 1957 or 1975. The Oil Shock did not play an important role for **d/y**. Over the most recent 30 years, **d/y**'s performance is again negative IS and OOS. The full-sample OOS performance is also again negative:

d/y	Recent 30 years	All years
IS $\overline{R}^2$	−5.52%	0.91%
OOS $\overline{R}^2$	−20.79%	−1.93%

Earnings Price Ratio : Figure 1 shows that **e/p** had inferior performance until WW II, and superior performance from WW II to the late 1970s. After the Oil Shock, it had generally nondescript performance (with the exception of the late 1990s and early 2000s). Its best sample period was 1943 to 2002. 2003 and 2004 were bad years for this model. Over the most recent 30 years, **e/p**'s performance is again negative IS and OOS. The full-sample OOS performance is negative too.

e/p	Recent 30 years	All years
IS $\overline{R}^2$	−2.08%	1.08%
OOS $\overline{R}^2$	−5.98%	−1.78%

Table 1 shows that these three price ratios are not statistically significant IS at the 90% level. However, some disagreement in the literature can be explained by differences in the estimation period.[7]

Other Variables : The remaining plots in Figure 1 and the remaining IS insignificant models in Table 1 show that **d/e**, **dfy**, and **infl** essentially never had significantly positive OOS periods, and that **svar** had a huge drop in OOS performance from 1930 to 1933. Other variables (that are

[7] For example, the final lines in Table 1 show that **d/y** and **e/p** had positive and statistically significant IS performance at the 90% level if all data prior to 1927 is ignored. Nevertheless, Table 1 also shows that the OOS-$\overline{R}^2$ performance remains negative for both of these. Moreover, when the data begins in 1927 and the forecast begins in 1947 (another popular period choice), we find

(Data Begins in 1927)	**e/p**		**d/y**	
(Forecast Begins in 1947)	Recent	All	Recent	All
IS $\overline{R}^2$	−3.83%	3.20%	−5.20%	2.71%
OOS $\overline{R}^2$	−13.58%	3.41%	−28.05%	−16.65%

Finally, and again not reported in the table, another choice of estimation period can also make a difference. The three price models lost statistical significance over the full sample only in the 1990s. This is not because the IS-ΔRMSE decreased further in the 1990s, but because the 1991–2005 prediction errors were more volatile, which raised the standard errors of point estimates.

IS insignificant) often had good sample performance early on, ending somewhere between the Oil Shock and the mid-1980s, followed by poor performance over the most recent three decades. The plots also show that it was generally not just the late 1990s that invalidated them, unlike the case with the aforementioned price ratio models.

In sum, 12 models had insignificant IS full-period performance and, not surprisingly, these models generally did not offer good OOS performance.

3.2 In-sample significant models

Five models were significant IS (**b/m**, **i/k**, **ntis**, **eqis**, and **all**) at least at the 10% two-sided level. Table 1 contains more details for these variables, such as the IS performance during the OOS period, and a power statistic. Together with the plots in Figure 2, this information helps the reader to judge the stability of the models—whether poor OOS performance is driven by less accurately estimated parameters (pointing to lower power), and/or by the fact that the model fails IS and/or OOS during the OOS sample period (pointing to a spurious model).

Book-to-market ratio: b/m is statistically significant at the 6% level IS. Figure 2 shows that it had excellent IS and OOS predictive performance right until the Oil Shock. Both its IS and OOS performance were poor from 1975 to 2000, and the recovery in 2000–2002 was not enough to gain back the 1997–2000 performance. Thus, the **b/m** model has negative performance over the most recent three decades, both IS and OOS.

b/m	Recent 30 years	All years
IS $\overline{R}^2$	−12.37%	3.20%
OOS $\overline{R}^2$	−29.31%	−1.72%

Over the entire sample period, the OOS performance is negative, too. The "IS for OOS" $\overline{R}^2$ in Table 1 shows how dependent **b/m**'s performance is on the first 20 years of the sample. The IS $\overline{R}^2$ is −7.29% for the 1965–2005 period. The comparable OOS $\overline{R}^2$ even reaches −12.71%.

As with other models, **b/m**'s lack of OOS significance is not just a matter of low test power. Table 1 shows that in the OOS prediction beginning in 1941, under the simulation of a stable model, the OOS statistic came out *statistically significantly* positive in 67%[8] of our (stable-model) simulations in which the IS regression was significant. Not reported in the table, positive performance (significant or insignificant) occurred in 78% of our

[8] The 42% applies to all simulation draws. It is the equivalent of the experiment conducted in some other articles. However, because OOS performance is relevant only when the IS performance is significant, this is the wrong measure of power.

simulations. A performance as negative as the observed ΔRMSE of −0.01 occurred in *none* of the simulations.

Investment-capital ratio : i/k is statistically significant IS at the 5% level. Figure 2 shows that, like **b/m**, it performed well only in the first half of its sample, both IS and OOS. About half of its performance, both IS and OOS, occurs during the Oil Shock. Over the most recent 30 years, **i/k** has underperformed:

i/k	Recent 30 years	All years
IS $\overline{R}^2$	−8.09%	6.63%
OOS $\overline{R}^2$	−18.02%	−1.77%

Corporate Issuing Activity : Recall that **ntis** measures equity issuing and repurchasing (plus dividends) relative to the price level; **eqis** measures equity issuing relative to debt issuing. Figure 2 shows that both variables had superior IS performance in the early 1930s, a part of the sample that is not part of the OOS period. **eqis** continues good performance into the late 1930s but gives back the extra gains immediately thereafter. In the OOS period, there is one stark difference between the two variables: **eqis** had superior performance during the Oil Shock, both IS and OOS. It is this performance that makes **eqis** the only variable that had statistically significant OOS performance in the annual data. In other periods, neither variable had superior performance during the OOS period.

Both variables underperformed over the most recent 30 years

	ntis		**eqis**	
	Recent 30 years	All years	Recent 30 years	All years
IS $\overline{R}^2$	−5.14%	8.15%	−10.36%	9.15%
OOS $\overline{R}^2$	−8.63%	−5.07%	−15.33%	2.04%

The plot can also help explain dueling perspectives about **eqis** between Butler, Grullon, and Weston (2005) and Baker, Taliaferro, and Wurgler (2004). One part of their disagreement is whether **eqis**'s performance is just random underperformance in sampled observations. Of course, some good years are expected to occur in any regression. Yet **eqis**'s superior performance may not have been so random, because it (i) occurred in consecutive years, and (ii) in response to the Oil Shock events that are often considered to have been exogenous, unforecastable, and unusual. Butler, Grullon, and Weston (2005) also end their data in 2002, while Baker, Taliaferro, and Wurgler (2004) refer to our earlier draft and to Rapach and Wohar (2006), which end in 2003 and 1999, respectively. Our figure shows that small variations in the final year choice can make a

difference in whether **eqis** turns out significant or not. In any case, both articles have good points. We agree with Butler, Grullon, and Weston (2005) that **eqis** would not have been a profitable and reliable predictor for an external investor, especially over the most recent 30 years. But we also agree with Baker, Taliaferro, and Wurgler (2004) that conceptually, it is not the OOS performance, but the IS performance that matters in the sense in which Baker and Wurgler (2000) were proposing **eqis**—not as a third-party predictor, but as documentary evidence of the fund-raising behavior of corporations. Corporations did repurchase profitably in the Great Depression and the Oil Shock era (though not in the "bubble period" collapse of 2001–2002).

all The final model with IS significance is the kitchen sink regression. It had high IS significance, but exceptionally poor OOS performance.

3.3 Time-changing models

caya and **ms** have no IS analogs, because the models themselves are constantly changing.

Consumption-Wealth-Income : Lettau and Ludvigson (2001) construct their **cay** proxy assuming that agents have some *ex-post* information. The experiment their study calls OOS is unusual: their representative agent still retains knowledge of the model's full-sample CAY-*construction* coefficients. It is OOS only in that the agent does not have knowledge of the *predictive* coefficient and thus has to update it on a running basis. We call the Lettau and Ludvigson (2001) variable **cayp**. We also construct **caya**, which represents a more genuine OOS experiment, in which investors are not assumed to have advance knowledge of the **cay** construction estimation coefficients.

Figure 2 shows that **cayp** had superior performance until the Oil Shock, and nondescript performance thereafter. It also benefited greatly from its performance during the Oil Shock itself.

cay	Recent 30 years	All years
Some *ex-post* knowledge, **cayp** IS $\overline{R}^2$	10.52%	15.72%
Some *ex-post* knowledge, **cayp** OOS $\overline{R}^2$	7.60%	16.78%
No advance knowledge, **caya** OOS $\overline{R}^2$	−12.39%	−4.33%

The full-sample **cayp** result confirms the findings in Lettau and Ludvigson (2001). **cayp** outperforms the benchmark OOS RMSE by 1.61% per annum. It is stable and its OOS performance is almost identical to its IS performance. In contrast to **cayp**, **caya** has had no superior OOS performance, either over the entire sample period or the most recent years. In fact, without advance knowledge, **caya** had the worst OOS $\overline{R}^2$ performance among our single variable models.

Model Selection : Finally, **ms** fails with a pattern similar to earlier variables—good performance until 1976, bad performance thereafter.

	Recent	All
ms	30 years	years
IS $\overline{R}^2$		
OOS $\overline{R}^2$	−43.40%	−22.50%

Conclusion : There were a number of periods with sharp stock market changes, such as the Great Depression of 1929–1933 (in which the S&P500 dropped from 24.35 at the end of 1928 to 6.89 at the end of 1932) and the "bubble period" from 1999–2001 (with its subsequent collapse). However, it is the Oil Shock recession of 1973–1975, in which the S&P500 dropped from 108.29 in October 1973 to 63.54 in September 1974—and its recovery back to 95.19 in June 1975—that stands out. Many models depend on it for their apparent forecasting ability, often both IS and OOS. (And none performs well thereafter.) Still, we caution against overreading or underreading this evidence. In favor of discounting this period, the observed source of significance seems unusual, because the important years are consecutive observations during an unusual period. (They do not appear to be merely independent draws.) In favor of not discounting this period, we do not know how one would identify these special multiyear periods ahead of time, except through a model. Thus, good prediction during such a large shock should not be automatically discounted. More importantly and less ambiguously, no model seems to have performed well since—that is, over the last 30 years.

In sum, on an annual prediction basis, there is no single variable that meets all of our four suggested investment criteria (IS significance, OOS performance, reliance not just on some outliers, and good positive performance over the last three decades.) Most models fail on all four criteria.

4. Five-yearly Prediction

Some models may predict long-term returns better than short-term returns. Unfortunately, we do not have many years to explore five-year predictions thoroughly, and there are difficult econometric issues arising from data overlap. Therefore, we only briefly describe some preliminary and perhaps naive findings. (See, e.g., Boudoukh, Richardson and Whitelaw (2005) and Lamoureux and Zhou (1996) for more detailed treatments.) Table 2 repeats Table 1 with five-year returns. As before, we bootstrap all critical significance levels. This is especially important here, because the observations are overlapping and the asymptotic critical values are not available.

Table 2 shows that there are four models that are significant IS over the entire sample period: **ntis**, **d/p**, **i/k**, and **all**. **ntis** and **i/k** were also significant

Table 2
Forecasts at 5-year frequency This table is identical to Table 1, except that we predict overlapping 5-yearly equity premia, rather than annual equity premia

			Full sample									1927–2005
				Forecasts begin 20 years after sample				Forecasts begin 1965				Sample
			IS	IS for	OOS			IS for	OOS			IS
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	$\overline{R}^2$
Full Sample, Not Significant IS												
ltr	Long term return	1926–2005	−1.36		−7.40	−1.10			−18.92	−2.72		−1.39
dfr	Default return spread	1926–2005	−1.36		−5.71	−0.77			−4.01	−0.25		−1.36
infl	Inflation	1919–2005	−1.21		−11.25	−1.70			−7.34	−0.85		−1.21
lty	Long term yield	1919–2005	−0.15		−122.13	−17.41			−72.47	−10.96		−0.30
svar	Stock variance	1885–2005	0.33		−79.33	−13.31			−2.37	0.03		−0.84
d/e	Dividend payout ratio	1872–2005	0.66		−4.87	−0.76			−0.64	0.31		1.64
dfy	Default yield spread	1919–2005	3.54		−59.33	−9.18			4.97*	+1.38		0.94
tbl	Treasury-bill rate	1920–2005	3.83		−17.66	−2.78			−30.19	−4.70		4.91
d/y	Dividend yield	1872–2005	6.04		−4.45	−0.68			−16.84	−2.19		**14.99***
e/p	Earning price ratio	1872–2005	6.24		−1.04	−0.03			−3.03	−0.07		**14.96***
tms	Term spread	1920–2005	7.84		−26.52	−4.24			**10.46****	+2.44		**12.47***
eqis	Pct equity issuing	1927–2005	9.50		−2.35	−0.11			−5.75	−0.56		Same
b/m	Book to market	1921–2005	10.78		−13.06	−2.03			−46.34	−7.17		13.93
Full Sample, Significant IS												
ntis	Net equity expansion	1927–2005	**6.59***	−8.28	−3.46	−0.32	21(70)	1.49	−13.77	−1.92	21(67)	Same
d/p	Dividend price ratio	1872–2005	**10.24***	14.35	−1.19*	−0.06	21(69)	8.30	−26.09	−3.54	20(51)	**21.24****
i/k	Invstmnt capital ratio	1947–2005	**33.99*****	27.42	**12.99****	+3.39	22(78)			Same		Same
all	Kitchen sink	1927–2005	**41.48*****	43.29	−499.83	−45.47	−(−)	19.75	−442.08	−34.19	−(−)	Same

Table 2
(Continued)

			Full sample									1927–2005
				Forecasts begin 20 years after sample				Forecasts begin 1965				Sample
			IS	IS for	OOS			IS for	OOS			IS
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	$\overline{R}^2$
Full Sample, No IS Equivalent (caya, ms) or Ex-Post Information (cayp)												
cayp	Cnsmptn, wlth, incme	1945–2005	**36.05*****	63.11	**30.35*****	+7.50	- (-)			Same		Same
caya	Cnsmptn, Wlth, Incme	1945–2005	–	–	**9.10*****	+2.50	- (-)			Same		Same
ms	Model Selection	1927–2005	–	–	−14465.67	−408.06	- (-)	–	−122.89	−18.03	−(−)	Same
1927-2005 Sample, Significant IS												Full Sample
tms	Term Spread	1927–2005	**12.47***					23.24	**12.59****	+2.77	11(65)	7.84
e/p	Earning Price Ratio	1927–2005	**14.96***					−4.04	−15.33	−2.18	28(65)	6.24
d/y	Dividend Yield	1927–2005	**14.99***					6.16	−9.47	−1.19	22(72)	6.04
d/p	Dividend Price Ratio	1927–2005	**21.24****					4.28	−12.69	−1.74	29(61)	**10.24***

in the annual data (Table 1). Two more variables, **d/y** and **tms**, are IS significant if no data prior to 1927 is used.

Dividend Price Ratio : d/p had negative performance OOS regardless of period.

Term Spread : tms is significant IS only if the data begins in 1927 rather than 1921. An unreported plot shows that **tms** performed well from 1968 to 1979, poorly from 1979 to 1986, and then well again from 1986 to 2005. Indeed, its better years occur in the OOS period, with an IS $\overline{R}^2$ of 23.54% from 1965 to 2005. This was sufficient to permit it to turn in a superior OOS ΔRMSE performance of 2.77% per five-years—a meaningful difference. On the negative side, **tms** has positive OOS performance *only* if forecasting begins in 1965. Using 1927–2005 data and starting forecasts in 1947, the OOS ΔRMSE and $\overline{R}^2$ are negative.

The Kitchen Sink : all again turned in exceptionally poor OOS performance.

Model selection (**ms**) and **caya** again have no IS analogs. **ms** had the worst predictive performance observed in this paper. **caya** had good OOS performance of 2.50% per five-year period. Similarly, the investment-capital ratio, **i/k**, had both positive IS and OOS performance, and both over the most recent three decades as well as over the full sample (where it was also statistically significant).

i/k	Recent 30 years	All years
IS $\overline{R}^2$	30.60%	33.99%
OOS $\overline{R}^2$	28.00%	12.99%

i/k's performance is driven by its ability to predict the 2000 crash. In 1997, it had already turned negative on its 1998–2002 equity premium prediction, thus predicting the 2000 collapse, while the unconditional benchmark prediction continued with its 30% plus predictions:

Forecast made in	For years	Actual EqPm	Forecast Unc.	Forecast **i/k**	Forecast made in	For years	Actual EqPm	Forecast Unc.	Forecast **i/k**
1995	1996–2000	0.58	0.30	0.22	1998	1999–2003	−0.19	0.33	−0.09
1996	1997–2001	0.27	0.31	0.09	1999	2000–2004	−0.25	0.34	−0.07
1997	1998–2002	−0.23	0.31	−0.01	2000	2001–2005	−0.08	0.34	−0.06

This model (and perhaps **caya**) seem promising. We hesitate to endorse them further only because our inference is based on a small number of observations, and because statistical significance with overlapping multiyear returns raises a set of issues that we can only tangentially address. We hope more data will allow researchers to explore these models in more detail.

5. Monthly Prediction and Campbell–Thompson

Table 3 describes the performance of models predicting monthly equity premia. It also addresses a number of points brought up by Campbell and Thompson (2005), henceforth CT. We do not have dividend data prior to 1927, and thus no reliable equity premium data before then. This is why even our the estimation period begins only in 1927.

5.1 In-sample performance

Table 3 presents the performance of monthly predictions both IS and OOS. The first data column shows the IS performance when the predicted variable is logged (as in the rest of the article). Eight out of eighteen models are IS significant at the 90% level, seven at the 95% level. Because CT use simple rather than log equity premia, the remaining data columns follow their convention. This generally improves the predictive power of most models, and the fourth column (by which rows are sorted) shows that three more models turn in statistically significant IS.[9]

CT argue that a reasonable investor would not have used a model to forecast a negative equity premium. Therefore, they suggest truncation of such predictions at zero. In a sense, this injects caution into the models themselves, a point we agree with. Because there were high equity premium realizations especially in the 1980s and 1990s, a time when many models were bearish, this constraint can improve performance. Of course, it also transforms formerly linear models into nonlinear models, which are generally not the subject of our paper. CT do *not* truncate predictions in their IS regressions, but there is no reason not to do so. Therefore, the fifth column shows a revised IS $\overline{R}^2$ statistic. Some models now perform better, some perform worse.

5.2 Out-of-sample prediction performance

The remaining columns explore the OOS performance. The sixth column shows that without further manipulation, **eqis** is the only model with both superior IS ($\overline{R}^2 = 0.82\%$ and 0.80%) and OOS ($\overline{R}^2 = 0.14\%$) untruncated performance. The term-spread, **tms**, has OOS performance that is even better ($\overline{R}^2 = 0.22\%$), but it just misses statistical significance IS at the 90% level. **infl** has marginally good OOS performance, but poor IS performance. All other models have negative IS or OOS untruncated $\overline{R}^2$.

The remaining columns show model performance when we implement the Campbell and Thompson (2005) suggestions. The seventh column describes the frequency of truncation of negative equity premium

[9] Geert Bekaert pointed out to us that if returns are truly log-normal, part of their increased explanatory power could be due to the ability of these variables to forecast volatility.

Table 3
Forecasts at monthly frequency using Campbell and Thompson (2005) procedure

Refer to Table 1 for basic explanations. This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for equity premium forecasts at the monthly frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. The data period is December 1927 to December 2004, except for csp (May 1937 to December 2002) and cay3 (December 1951 to December 2004). Critical values of all statistics are obtained empirically from bootstrapped distributions, except for cay3 model where they are obtained from McCracken (2004). The resulting significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively. They are two-sided for IS model significance, and one-sided for OOS superior model performance. The first data column is the IS $\overline{R}^2$ when returns are logged, as they are in our other tables. The remaining columns are based on predicting simple returns for correspondence with Campbell and Thompson (2005). Certainty Equivalence (CEV) gains are based on the utility of an optimizer with a risk-aversion coefficient of $\gamma = 3$ who trades based on unconditional forecast and conditional forecast. Equity positions are winsorized at 150% ($w = w_{\max}$). At this risk-aversion, the base CEV are 82bp for a market-timer based on the unconditional forecast, 79bp for the market, and 40bp for the risk-free rate. "T" means "truncated" to avoid a negative equity premium prediction. "U" means unconditional, that is, to avoid a forecast that is based on a coefficient that is inverse to what the theory predicts. A superscript *h* denotes high trading turnover of about 10%/month more than the trading strategy based on unconditional forecasts.

		Log returns	Simple returns									
			IS		OOS	Campbell and Thompson (2005) OOS						
	Variable	IS $\overline{R}^2$	$\overline{R}^2$	$\overline{R}^2$ T	$\overline{R}^2$	Frcst= T	U	$\overline{R}^2$ TU	ΔRMSE TU	$w = w_{\max}$	ΔCEV	Fig
d/e	Dividend payout ratio	0.02	−0.10	−0.10	−0.70	0.0	7.9	−0.69	−0.0114	57.7	−0.01	
svar	Stock variance	−0.09	−0.07	−0.07	−0.79	0.0	0.0	−0.79	−0.0134	35.4	−0.04	
dfr	Default return spread	−0.02	−0.07	−0.08	−0.37	0.0	20.9	−0.29	−0.0030	44.9	0.01	
lty	Long term yield	−0.03	0.02	0.02	−0.80	34.1	0.0	**0.26****	+0.0085	19.5	0.06	
ltr	Long term return	0.04	0.07	0.08	−0.63	3.0	38.2	**0.11****	+0.0053	51.2[h]	0.06	
infl	Inflation	−0.01	0.14	−0.05	**0.01***	1.3	0.0	**0.07****	+0.0045	43.5[h]	0.04	
tms	Term spread	0.12	0.18	0.20	**0.22****	3.7	0.0	**0.21****	+0.0073	59.3	0.14	*F3.G*
tbl	Treasury-bill rate	0.10	**0.20***	0.15	−0.08*	23.1	0.0	**0.25****	+0.0081	16.4	0.10	*F3.F*
dfy	Default yield spread	−0.06	**0.28***	0.28	−0.56	4.0	0.0	−0.49	−0.0071	27.3	−0.08	
d/p	Dividend price ratio	0.12	**0.33***	0.29	−0.30	32.3	0.0	**0.17***	+0.0066	16.1	−0.10	*F3.E*
d/y	Dividend yield	**0.22***	**0.47****	0.45	−1.12	54.2	0.0	−0.04*	+0.0023	16.4	−0.14	
e/p	Earning price ratio	**0.51****	**0.54****	0.45	−1.04	18.1	0.0	−1.03	−0.0183	34.4	−0.04	
eqis	Pct equity issuing	**0.82*****	**0.80*****	0.59	**0.14****	6.7	0.0	**0.30*****	+0.0093	55.8	0.14	*F3.D*
b/m	Book to market	**0.45****	**0.81*****	0.88	−3.28	44.3	0.0	−2.23	−0.0432	31.3	−0.22	
e^{10}/p	Earning(10Y) price ratio	**0.46****	**0.86*****	0.96	−2.21	52.4	0.0	−0.48	−0.0071	15.4	−0.13	
csp	Cross-sectional prem	**0.92*****	**0.99*****	0.93	−0.94	44.7	0.0	**0.15****	+0.0072	13.5	0.06	*F3.B*
ntis	Net equity expansion	**0.94*****	**1.02*****	0.88	−0.16	0.4	0.0	−0.16	−0.0003	57.4	0.02	*F3.C*
cay3	Cnsmptn, wlth, incme	**1.88*****	**1.87*****	1.57	−2.05	44.7	0.0	−0.34*	+0.0088	13.2	0.06	*F3.A*

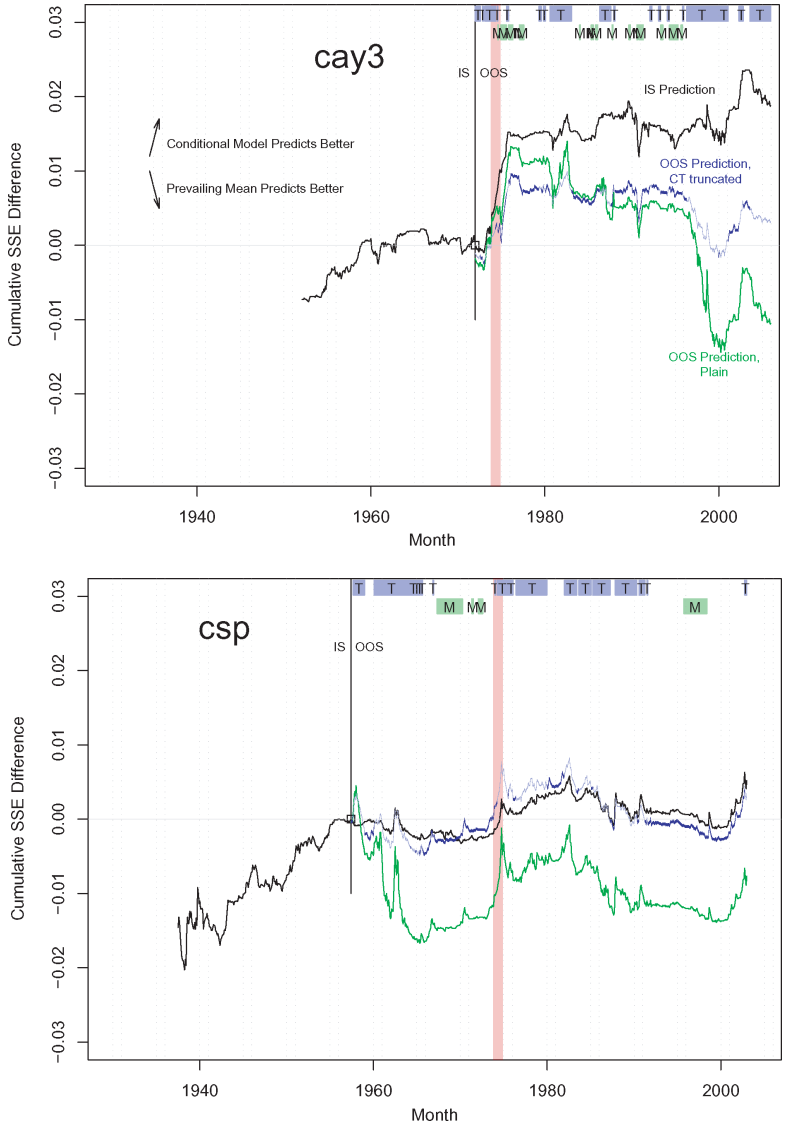

Figure 3
Monthly performance of in-sample significant predictors
Explanation: These figures are the analogs of Figures 1 and 2, plotting the IS and OOS performance of the named model. However, they use monthly data. The IS performance is in black. The Campbell-Thompson (2005) (CT) OOS model performance is plotted in blue, the plain OOS model performance is plotted in green. The top bars ("T") indicate truncation of the equity prediction at 0, inducing the CT investor to hold the risk-free security. (This also lightens the shade of blue in the CT line.) The lower bars ("M") indicate when the CT risk-averse investor would purchase equities worth 150% of his wealth, the maximum permitted. The Oil Shock (Nov 1973 to Mar 1975) is marked by a red vertical line.

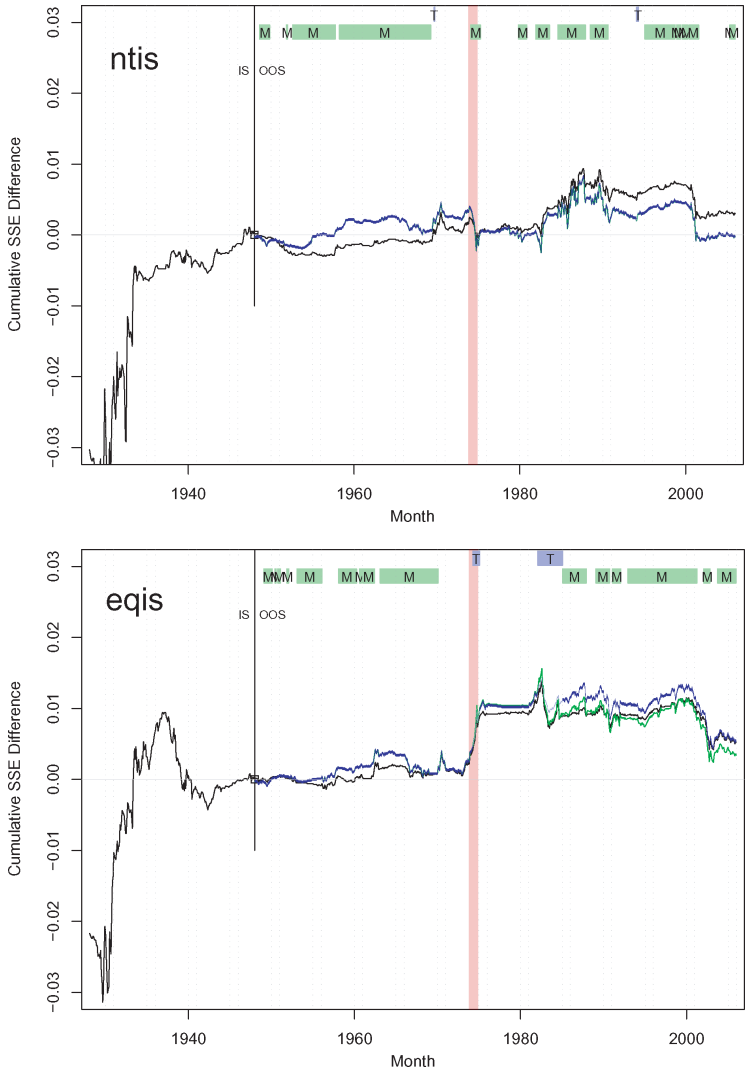

Figure 3 Continued

predictions. For example, **d/y**'s equity premium predictions are truncated to zero in 54.2% of all months; **csp**'s predictions are truncated in 44.7% of all months. Truncation is a very effective constraint.

CT also suggest using the unconditional model if the theory offers one coefficient sign and the estimation comes up with the opposite sign. For some variables, such as the dividend ratios, this is easy. For other models, it is not clear what the appropriate sign of the coefficient would be. In any case, this matters little in our data set. The eighth column shows that the

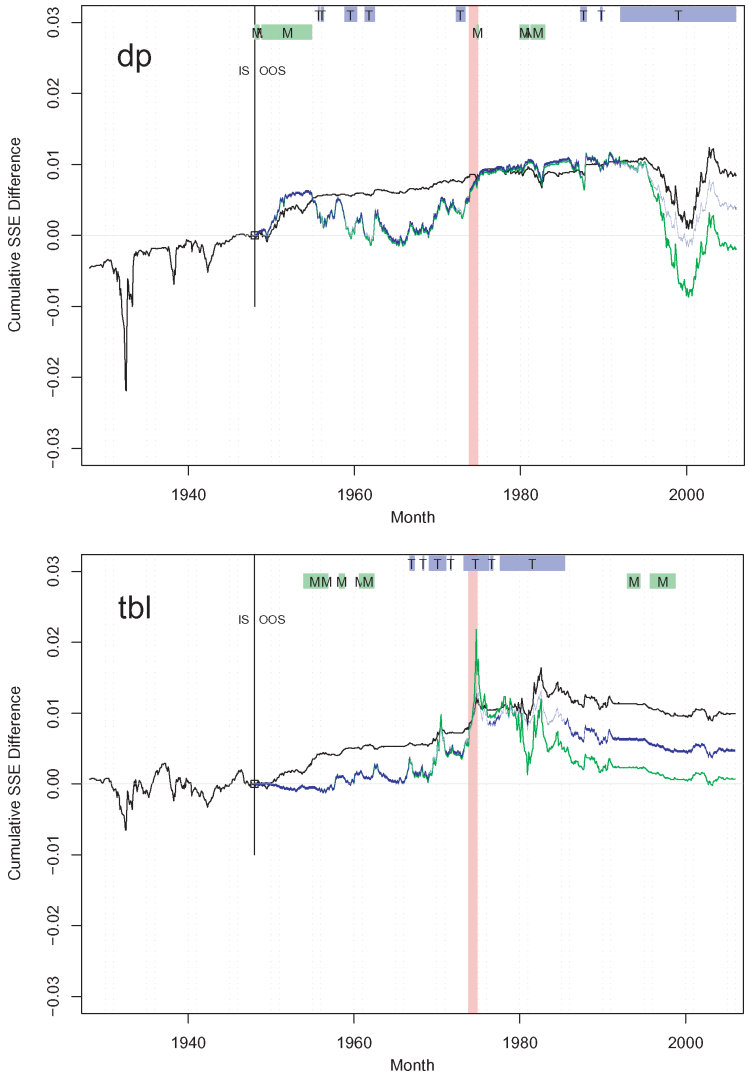

Figure 3 Continued

coefficient sign constraint matters only for **dfr** and **ltr** (and mildly for d/e). None of these three models has IS performance high enough to make this worthwhile to explore further.

The ninth and tenth columns, $\overline{R}^2_{\text{TU}}$ and $\Delta\text{RMSE}_{\text{TU}}$, show the effect of the CT truncations on OOS prediction. For many models, the performance improves. Nevertheless, the OOS $\overline{R}^2$'s remain generally much lower than their IS equivalents. Some models have positive ΔRMSE but negative

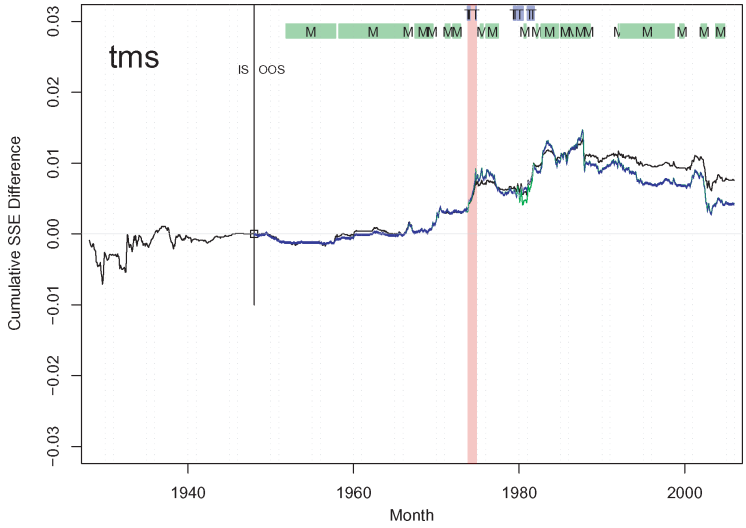

Figure 3 Continued

OOS $\overline{R}^2$. This reflects the number of degrees of freedom: even though we have between 400 and 800 data months, the plain ΔRMSE and R^2 are often so small that the $\overline{R}^2$ turns negative. For example, even with over 400 months of data, the loss of three degrees of freedom is enough for **cay3** to render a positive ΔRMSE of 0.0088 (equivalent to an unreported unadjusted R^2 of 0.0040) into a negative adjusted R^2 of -0.0034.

Even after these truncations, ten of the models that had negative plain OOS $\overline{R}^2$'s still have negative CT OOS $\overline{R}^2$'s. Among the eleven IS significant models, seven (**cay3**, **ntis**, **e^{10}/p**, **b/m**, **e/p**, **d/y**, and **dfy**) have negative OOS $\overline{R}^2$ performance even after the truncation. Three of the models (**lty**, **ltr**, and **infl**) that benefit from the OOS truncation are not close to statistical significance IS, and thus can be ignored. All in all, this leaves four models that are both OOS and IS positive and significant: csp, **eqis**, **d/p**, **tbl**, plus possibly **tms** (which is just barely not IS significant). We investigate these models further below.

5.3 OOS utility performance of a trading strategy

Like Brennan and Xia (2004), CT also propose to evaluate the OOS usefulness of models based on the certainty equivalence (CEV) measure of a trading strategy. Specifically, they posit a power-utility investor with an assumed risk-aversion parameter, γ, of three. This allows a conditional model to contribute to an investment strategy not just by increasing the mean trading performance, but also by reducing the variance. (Breen,

Glosten and Jagannathan (1989) have shown this to be a potentially important factor.)

Although the focus of our article is on mean prediction, we know of no better procedure to judge the economic significance of forecasting models, and therefore follow their suggestion here. To prevent extreme investments, there is a 150% maximum equity investment. A positive investment weight is guaranteed by the truncation of equity premium predictions at zero.

CT show that even a small improvement in ΔRMSE by a model over the unconditional benchmark can translate into CEV gains that are ten times as large.[10] We can confirm this—and almost to a fault. **cay3** offers 6.1bp/month performance, even though it had a negative $\overline{R}^2$. Column 12 also shows that even models that have a negative OOS ΔRMSE (not just a negative $\overline{R}^2$), like **dfr**, can produce positive gains in CEV. This is because the risk-aversion parameter γ of 3 is low enough to favor equity-tilted strategies. Put differently, some strategy CEV gains are due to the fact that the risky equity investment was a better choice than the risk-free rate in our data. (This applies not only to strategies based on the conditional models, but also to the strategy based on the unconditional mean.) An alternative utility specification that raises the risk-aversion coefficient to 7.48 would have left an investor indifferent between the risk-free and the equity investments. Briefly considering this parameter can help judge the role of equity bias in a strategy; it does seem to matter for the **eqis** and **tms** models, as explained below.

In order, among the IS reasonably significant models, those providing positive CEV gains were **tms** (14bp/month), **eqis** (14bp/month), **tbl** (10bp/month), **csp** (6bp/month), **cay3** (6bp/month), and **ntis** (2bp/month).

5.4 Details

We now look more closely at the set of variables with potentially appealing forecasting characteristics. **csp**, **eqis**, **tbl**, and **tms** have positive IS performance (either statistically significant or close to it), positive OOS $\overline{R}^2$ (truncated), and positive CEV gains. **cay3** and **ntis** have negative OOS $\overline{R}^2$, but very good IS performance and positive CEV gains. **d/p** has a negative CEV gain, but is positive IS and OOS $\overline{R}^2$. Thus, we describe these seven models in more detail (and with equivalent graphs):

[10] CT show in Equation (8) of their paper that the utility gain is roughly equal to OOS-R^2/γ. This magnification effect occurs only on the monthly horizon, because the difference between OOS-R^2 and the ΔRMSE scales with the square root of the forecasting horizon (for small ΔRMSE, OOS-$R^2 \approx 2\times\Delta\text{RMSE}/\text{StdDev}(R)$). That is, at a monthly frequency, the OOS-R^2 is about 43 times as large as ΔRMSE. On an annual prediction basis, this number drops from 43 to 12. An investor with a risk aversion of 10 would therefore consider the economic significance on annual investment horizon to be roughly the same as the ΔRMSE we consider. (We repeated the CT CEV equivalent at annual frequency to confirm this analysis.)

(i) **cay3:** The best CT performer is an alternative **cay** model that also appears in Lettau and Ludvigson (2005). It predicts the equity premium not with the linear **cay**, but with all three of its highly cointegrated ingredients up to date. We name this model **cay3**. In unreported analysis, we found that the **cay** model and **cay3** models are quite different. For most of the sample period, the unrestricted predictive regression coefficients of the **cay3** model wander far off their cointegration-restricted **cay** equivalents. The model may not be as well founded theoretically as the Lettau and Ludvigson (2001) **cay**, but if its components are known *ex-ante*, then **cay3** is fair game for prediction.

Table 3 shows that **cay3** has good performance IS, but only marginal performance OOS (a positive ΔRMSE, but a negative $\overline{R}^2$). It offers good CEV gains among the models considered, an extra 6.10 bp/month. The h superscript indicates that its trading strategy requires an extra 10% more trading turnover than the unconditional model. It also reaches the maximum permitted 150% equity investment in 13.2% of all months.

A first drawback is that the **cay3** model relies on data that may not be immediately available. Its components are publicly released by the Bureau of Economic Analysis about 1–2 months after the fact. Adding just one month delay to trading turns **cay3**'s performance negative:

	ΔRMSE	$\Delta\text{RMSE}_{\text{TU}}$	ΔCEV
Immediate availability (CT)	−2.88 bp	+0.88 bp	+6.10 bp
One month delayed	−5.10 bp	−1.62 bp	−11.82 bp
Two months delayed	−5.38 bp	−1.11 bp	−9.80 bp

A second drawback is visible in Figure 3. Like **caya** and **cayp**, much of **cay3**'s performance occurs around the Oil Shock (most of its OOS performance are between one-half and one-third of its IS performance). Even IS, **cay3** has not performed well for over 30 years now:

cay3 (CT)	Recent 30 years	All years
IS $\overline{R}^2$	−0.30%	1.87%
OOS $\overline{R}^2$	−1.60%	−0.34%

Finally, the figure shows that many of **cay3**'s recent equity premium forecasts have been negative and therefore truncated. And, therefore, the information in its current forecasts is limited.

(ii) **csp:** Table 3 shows that the relative valuations of high- over low-beta stocks had good IS and truncated OOS performance,

and offered a market timer 6.12 bp/month superior the CEV-equivalent performance. The plot in Figure 3 shows that **csp** had good performance from September 1965 to March 1980. It underperformed by just as much from about April 1980 to October 2000. In fact, from its first OOS prediction in April 1957 to August 2001, **csp**'s total net performance was zero even after the CT truncations, and both IS and OOS. All of **csp**'s superior OOS performance has occurred since mid-2001. Although it is commendable that it has performed well late rather than early, better performance over its first 45 years would have made us deem this variable more reliable.

The plot raises one other puzzle. The CT-truncated version performs better than the plain OLS version because it truncated the **csp** predictions from July 1957 through January 1963. These CT truncations are critically responsible for its superior OOS performance, but make no difference thereafter. It is the truncation treatment of these specific 66 months that would make an investor either believe in superior positive or inferior outright negative performance for **csp** (from August 2001 to December 2005). We do not understand why the particular 66 month period from 1957 to 1963 is so crucial.

Finally, the performance during the Oil Shock recession is not important for IS performance, but it is for the OOS performance. It can practically account for its entire OOS performance. Since the Oil Shock, **csp** has outperformed IS, but not OOS:

csp(CT)	Recent 30 years	All years
IS $\overline{R}^2$	0.33%	0.99%
OOS $\overline{R}^2$	−0.41%	0.15%

(iii) **ntis:** Net issuing activity had good IS performance, but a negative OOS $\overline{R}^2$. Its CEV gain is a tiny 1.53 bp/month. These 1.53 bp are likely to be offset by trading costs to turn over an additional 4.6% of the portfolio every month.[11] The strategy was very optimistic, reaching the maximum 150% investment constraint in 57.4% of all months. We do not report it in the table, but an investor with a higher 7.48 risk-aversion parameter, who would not have been so eager to highly lever herself into the market,

[11] Keim and Madhavan (1997) show that one typical roundtrip trade in large stocks for institutional investors would have conservatively cost around 38 bp from 1991–1993. Costs for other investors and earlier time-periods were higher. Futures trading costs are not easy to gage, but a typical contract for a notional amount of $250,000 costs around $10–$30. A 20% movement in the underlying index—about the annual volatility—would correspond to $50,000, which would come to around 5 bp.

would have experienced a negative CEV with an **ntis** optimized trading strategy. Finally, the plot shows that almost all of the **csp** model's IS power derives from its performance during the Great Depression. There was really only a very short window from 1982 to 1987 when **csp** could still perform well.

(iv) **eqis:** Equity Issuing Activity had good IS performance and a good OOS performance, and improved the CEV for an investor by a meaningful 13.67 bp/month. It, too, was an optimistic equity-aggressive strategy. With a $\gamma = 3$, trading based on this variable leads to the maximum permitted equity investment of 150% in 56% of all months. Not reported, with the higher risk-aversion coefficient of 7.48, that would leave an investor indifferent between bonds and stocks, the 13.67 bp/month gain would shrink to 8.74 bp/month.

As in the annual data, Figure 3 shows that **eqis**'s performance relies heavily on the good Oil Shock years. It has not performed well in the last 30 years.

eqis(CT)	Recent 30 years	All years
IS $\overline{R}^2$	−0.88%	0.80%
OOS $\overline{R}^2$	−1.00%	0.30%

(v) **d/p:** The dividend price ratio has good IS and OOS $\overline{R}^2$. (The OOS $\overline{R}^2$ is zero when predicting log premia.) An investor trading on **d/p** would have lost the CEV of 10 bp/month. (Not reported, a more risk-averse investor might have broken even.) The plot shows that **d/p** has not performed well over the last 30 years; **d/p** has predicted negative equity premia since January 1992.

d/p(CT)	Recent 30 years	All years
IS $\overline{R}^2$	−0.39%	0.33%
OOS $\overline{R}^2$	−1.09%	0.17%

(vi) **tbl:** The short rate is insignificant IS if we forecast log premia. If we forecast unlogged premia, it is statistically significant IS at the 9% level, although this declines further if we apply the CT truncation. In its favor, **tbl**'s full-sample CT-truncated performance is statistically significant OOS, and it offers a respectable 9.53 bp/month market timing advantage. The plot shows that this is again largely Oil Shock dependent. **tbl** has offered no advantage over the last thirty years.

tbl(CT)	Recent 30 years	All years
IS $\overline{R}^2$	−0.41%	0.20%
OOS $\overline{R}^2$	−1.06%	0.25%

(vii) **tms:** The term-spread has IS significance only at the 10.1% level. (With logged returns, this drops to the 14.5% level.) Nevertheless, **tms** had solid OOS performance, either with or without the CT truncation. As a consequence, its CEV gain was a respectable 14.40 bp/month. Not reported in the table, when compared to the CEV gain of an investor with a risk-aversion coefficient of 7.48, we learn that about half of this gain comes from the fact that the term-spread was equity heavy. (It reaches its maximum of 150% equity investment in 59.3% of all months.) The figure shows that TMS performed well in the period from 1970 to the mid-1980s, that TMS has underperformed since then, and that the Oil Shock gain was greater than the overall OOS sample performance of **tms**. Thus,

tms(CT)	Recent 30 years	All years
IS $\overline{R}^2$	−0.19%	0.18%
OOS $\overline{R}^2$	−0.81%	0.21%

b/m, **e/p**, $\mathbf{e^{10}/p}$, **d/y**, and **dfy** have negative OOS $\overline{R}^2$ and/or CT CEV gains, and so are not further considered. The remaining models have low or negative IS $\overline{R}^2$, and therefore should not be considered, either. Not reported, among the models that are IS insignificant, but OOS significant, none had positive performance from 1975 till today.

5.5 Comparing findings and perspectives

The numbers we report are slightly different from those in Campbell and Thompson (2005). In particular, they report **cay3** to have a ΔRMSE of 0.0356, more than the 0.0088 we report. This can be traced back to three equally important factors: they end their data 34 months earlier (in 2/2003), they begin their estimation one month later (1/1952), and they use an earlier version of the **cay** data from Martin Lettau's website. Differences in other variables are sometimes due to use of pre-1927 data (relying on price changes because returns are not available) for estimation though not prediction, while we exclude all pre-1927 data.

More importantly, our perspective is different from CT's. We believe that the data suggests not only that these models are not good enough for actual investing, but also that the models are not stable. Therefore, by and large, we consider even their IS significance to be dubious. Because they

fail stability diagnostics, we would recommend against their continued use. Still, we can agree with some points CT raise:

(i) One can reasonably truncate the models' predictions.
(ii) On shorter horizons, even a small predictive ΔRMSE difference can gain a risk-averse investor good CEV gains.
(iii) OOS performance should not be used for primary analysis.

We draw different conclusions from this last point. We view OOS performance not as a substitute but as a necessary complement to IS performance. We consider it to be an important regression *diagnostic*, and *if and only if* the model is significant IS. Consequently, we disagree with the CT analysis of the statistical power of OOS tests. In our view, because the OOS power matters only if the IS regression is statistically significant, the power of the OOS tests is conditional and thus much higher than suggested in CT, Cochrane (2005), and elsewhere. Of course, any additional diagnostic test can only reject a model—if an author is sure that the linear specification is correct, then not running the OOS test surely remains more powerful.

In judging the usefulness of these models, our article attaches more importance than CT to the following facts:

(i) Most models are not IS significant. That is, many variables in the academic literature no longer have IS significance (even at the 90% level). It is our perspective that this disqualifies them as forecasters for researchers without strong priors.
(ii) After three decades of poor performance, often even IS, one should further doubt the stability of most prediction models.
(iii) Even after the CT truncation, many models earn negative CEV gains.
(iv) What we call OOS performance is not truly OOS, because it still relies on the same data that was used to establish the models. (This is especially applicable to **eqis** and **csp**, which were only recently proposed.)
(v) For practical use, an investor would have had to have known *ex-ante* which of the models would have held up, and that none of the models had superior performance over the last three decades—in our opinion because the models are unstable.

We believe it is now best left to the reader to concur either with our or CT's perspective. (The data is posted on the website.)

6. Alternative Specifications

We now explore some other models and specifications that have been proposed as improvements over the simple regression specifications.

6.1 Longer-memory dividend and earnings ratios

Table 4 considers dividend–price ratios, earnings–price ratios, and dividend–earnings ratios with memory (which simply means that we consider sums of multiple year dividends or earnings in these ratios). The table is an excerpt from a complete set of one-year, five-year, and ten-year dividend–price ratios, earnings–price ratios, and dividend–earnings ratios. (That is, we tried all 90 possible model combinations.) The table contains *all* 27 IS significant specifications from our monthly regressions that begin forecasting in 1965, and from our annual and 5-yearly forecasts that begin forecasting either in 1902 or 1965.

Even though there were more combinations of dividend–earnings ratios than either dividend–price or earnings–price ratios, not a single dividend–earnings ratio turned out IS statistically significant. The reader can also see that out of our 27 IS–significant models, only 5 had OOS positive and statistically significant performance. (For 2 of these models, the OOS significance is modest, not even reaching the 95% significance level.) Unreported graphs show that none of these performed well over the last three decades. (We also leave it to the readers to decide whether they believe that real-world investors would have been able to choose the right five models for prediction, and to get out right after the Oil Shock.)

6.2 Different estimation methods to improve power for nonstationary independent variables

Stambaugh (1999) shows that predictive coefficients in small samples are biased if the independent variable is close to a random walk. Many of our variables have autoregressive coefficients above 0.5 on monthly frequency. Goyal and Welch (2003) show that **d/p** and **d/y**'s autocorrelations are not stable but themselves increase over the sample period, and similar patterns occur with other variables in our study. (The exceptions are **ntis**, **ltr**, and **dfy**.) Our previously reported statistics took stable positive autoregressive coefficients into account, because we bootstrapped for significance levels mimicking the IS autocorrelation of each independent variable.

However, one can use this information itself to design more powerful tests. Compared to the plain OLS techniques in our preceding tables, the Stambaugh coefficient correction is a more powerful test in nonasymptotic samples. There is also information that the autocorrelation is not constant for the dividend ratios, which we are ignoring in our current article. Goyal and Welch (2003) use rolling dividend–price ratio and dividend–growth autocorrelation estimates as instruments in their return predictions. This is model specific, and thus can only apply to one model, the dividend price ratio (**d/p**). In contrast, Lewellen (2004) and Campbell and Yogo (2006) introduce two further statistical corrections, extending Stambaugh (1999) and assuming different boundary behavior. This subsection, therefore, explores equity premium forecasts using these corrected coefficients.

Table 4
Significant forecasts using various d/p, e/p, and d/e Ratios

				IS	OOS	
	Variable	Data	Freq	$\overline{R}^2$	$\overline{R}^2$	ΔRMSE
e/p	Earning(1Y) price ratio	1927–2005	M 1965–	**0.54****	−1.20	−0.02
e^5/p	Earning(5Y) price ratio	1927–2005	M 1965–	**0.32***	−0.60	−0.01
e^{10}/p	Earning(10Y) price ratio	1927–2005	M 1965–	**0.49****	−0.83	−0.01
e^3/p	Earning(3Y) price ratio	1882–2005	A 1902–	**2.53****	−1.05*	−0.01
e^5/p	Earning(5Y) price ratio	1882–2005	A 1902–	**2.88****	−0.52*	+0.04
e^{10}/p	Earning(10Y) price ratio	1882–2005	A 1902–	**4.89****	**2.12****	+0.30
d^{10}/p	Dividend(3Y) price ratio	1882–2005	A 1902–	**1.85***	−1.53	−0.05
d^5/p	Dividend(5Y) price ratio	1882–2005	A 1902–	**2.48***	−0.54*	+0.04
d^{10}/p	Dividend(10Y) price ratio	1882–2005	A 1902–	**2.11***	−1.07*	−0.01
e^3/p	Earning(3Y) price ratio	1882–2005	A 1965–	**2.53****	−3.41	−0.06
e^5/p	Earning(5Y) price ratio	1882–2005	A 1965–	**2.88****	−5.01	−0.19
e^{10}/p	Earning(10Y) price ratio	1882–2005	A 1965–	**4.89****	−11.45	−0.66
d^3/p	Dividend(3Y) price ratio	1882–2005	A 1965–	**1.85***	−6.55	−0.30
d^5/p	Dividend(5Y) price ratio	1882–2005	A 1965–	**2.48***	−8.79	−0.47
d^{10}/p	Dividend(10Y) price ratio	1882–2005	A 1965–	**2.11***	−8.32	−0.43
e^3/p	Earning(3Y) price ratio	1882–2005	5Y 1902–	**11.35***	**3.46****	+0.89
e^5/p	Earning(5Y) price ratio	1882–2005	5Y 1902–	**16.16****	**4.76****	+1.16
e^{10}/p	Earning(10Y) price ratio	1882–2005	5Y 1902–	**16.47****	−2.85*	−0.37
d/p	Dividend(1Y) price ratio	1882–2005	5Y 1902–	**12.30***	−0.66*	+0.06
d^3/p	Dividend(3Y) price ratio	1882–2005	5Y 1902–	**13.11***	−2.02*	−0.21
d^5/p	Dividend(5Y) price ratio	1882–2005	5Y 1902–	**13.75***	−3.85*	−0.57
e^3/p	Earning(3Y) price ratio	1882–2005	5Y 1965–	**11.35***	−12.55	−1.56
e^5/p	Earning(5Y) price ratio	1882–2005	5Y 1965–	**16.16****	−21.16	−2.85
e^{10}/p	Earning(10Y) price ratio	1882–2005	5Y 1965–	**16.47****	−25.65	−3.51
d/p	Dividend(1Y) price ratio	1882–2005	5Y 1965–	**12.30***	−29.33	−4.03
d^3/p	Dividend(3Y) price ratio	1882–2005	5Y 1965–	**13.11***	−28.11	−3.86
d^5/p	Dividend(5Y) price ratio	1882–2005	5Y 1965–	**13.75***	−30.71	−4.23

Refer to Table 1 for basic explanations. The table reports only those combinations of **d/p**, **e/p**, and **d/e** that were found to predict equity premia significantly in-sample. This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at various frequencies. Variables are explained in Section 2. All ΔRMSE numbers are in percent per frequency corresponding to the column entitled 'Freq'. The 'Freq' column also gives the first year of forecast.

A star next to OOS-$\overline{R}^2$ is based on the MSE-F-statistic by McCracken (2004), which tests for equal MSE of the unconditional forecast and the conditional forecast. One-sided critical values of MSE statistics are obtained empirically from bootstrapped distributions. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

In Table 5, we predict with Stambaugh and Lewellen corrected coefficients. Both methods break the link between $\overline{R}^2$ (which is maximized by OLS) and statistical significance. The Lewellen coefficient is often dramatically different from the OLS coefficients, resulting in negative $\overline{R}^2$, even among its IS significant variable estimations. However, it is also tremendously powerful. Given our bootstrapped critical rejection levels under the NULL hypothesis, this technique is able to identify eight (rather

Table 5
Forecasts at monthly frequency with alternative procedures and total returns
Refer to Table 1 for basic explanations.Columns under the heading "OLS" are unadjusted betas, columns under the heading "Stambaugh" correct for betas following Stambaugh (1999), and columns under the heading 'Lewellen' correct for betas following Lewellen (2004). ρ under the column OLS gives the autoregressive coefficient of the variable over the entire sample period (the variables are sorted in descending order of ρ).

			OLS				Stambaugh			Lewellen		
			IS		OOS		IS	OOS		IS	OOS	
	Variable	Data	ρ	$\overline{R}^2$	$\overline{R}^2$	Power	$\overline{R}^2$	$\overline{R}^2$	Power	$\overline{R}^2$	$\overline{R}^2$	Power
d/e	Dividend payout ratio	192701–200512	0.9989	0.01	−2.02	15 (70)	0.01	−2.11	15 (69)	0.01	−2.05	15 (69)
lty	Long term yield	192701–200512	0.9963	−0.01	−1.15	9 (68)	−0.01	−1.71	9 (68)	−0.01	−1.05	10 (67)
d/y	Dividend yield	192701–200512	0.9929	**0.25***	−0.40	33 (71)	**0.25***	−0.36	33 (71)	**0.25***	−0.26	32 (72)
d/p	Dividend price ratio	192701–200512	0.9927	0.15	−0.15	29 (56)	0.05	−0.31	26 (69)	−0.15**	−1.03	3 (3)
tbl	Treasury-bill rate	192701–200512	0.9922	0.11	−0.18	19 (69)	0.11	−0.33	20 (69)	0.11	−0.27	20 (68)
e/p	Earning price ratio	192701–200512	0.9879	**0.54****	−1.21	56 (64)	**0.48****	−0.54	59 (73)	**0.02*****	−0.01**	41 (41)
b/m	Book to market	192701–200512	0.9843	**0.40****	−2.45	48 (65)	**0.36****	−1.61	48 (71)	−0.14**	−0.31	19 (19)
csp	Cross-Sectional prem	193705–200212	0.9788	**0.92*****	**0.70*****	65 (80)	**0.92*****	**0.70*****	65 (80)	**0.91*****	**0.71*****	65 (81)
dfy	Default yield spread	192701–200512	0.9763	−0.07	−0.14	9 (59)	−0.07	−0.33	8 (59)	−0.15	−0.71	5 (43)
ntis	Net equity expansion	192701–200512	0.9680	**0.75*****	−0.28	59 (76)	**0.75*****	−0.29	59 (76)	**0.74*****	−0.38	58 (74)
tms	Term spread	192701–200512	0.9566	0.07	**0.09***	21 (66)	0.07	**0.07***	20 (65)	0.07	**0.06***	20 (65)
svar	Stock variance	192701–200512	0.6008	−0.08	−0.34	7 (53)	−0.08	−0.34	7 (53)	−1.66**	−0.63	7 (7)
infl	Inflation	192701–200512	0.5513	−0.00	−0.07	14 (62)	−0.00	−0.07	14 (62)	−0.03	−0.13	15 (56)
ltr	Long term return	192701–200512	0.0532	0.04	−0.49	18 (62)	0.04	−0.48	18 (62)	−1.55***	−6.41	12 (12)
dfr	Default return spread	192701–200512	−0.1996	−0.02	−0.30	12 (61)	−0.02	−0.30	12 (61)	−1.32*	−2.64	10 (38)

than just three) ALTERNATIVE models as different from the NULL. In six of them, it even imputes significance in each and every one of our 10,000 bootstraps!

Unfortunately, neither the Stambaugh nor the Lewellen technique manages to improve OOS prediction. Of all models, only the **e/p** ratio in the Lewellen specification seems to perform better with a positive ΔRMSE. However, like other variables, it has not performed particularly well over the most recent 30 years—even though it has nonnegative OOS ΔRMSE (but not $\overline{R}^2$) performance over the last three decades.

e/p(Lewellen)	Recent 30 years	All years
IS $\overline{R}^2$	−0.16%	0.02%
OOS $\overline{R}^2$	−0.08%	−0.01%

6.3 Encompassing tests

Our next tests use encompassing predictions. A standard encompassing test is a hybrid of *ex-ante* OOS predictions and an *ex-post* optimal convex combination of unconditional forecast and conditional forecast. A parameter λ gives the *ex-post* weight on the conditional forecast for the optimal forecast that minimizes the *ex-post* MSE. The ENC statistic in Equation (7) can be regarded as a test statistic for λ. If λ is between 0 and 1, we can think of the combination model as a "shrinkage" estimator. It produces an optimal combination OOS forecast error, which we denote ΔRMSE*. However, investors would not have known the optimal *ex-post* λ. This means that they would have computed λ on the basis of the best predictive up-to-date combination of the two OOS model (NULL and ALTERNATIVE), and then would have used this λ to forecast one month ahead. We denote the relative OOS forecast error of this rolling λ procedure as ΔRMSE$^{\star r}$.[12]

Table 6 shows the results of encompassing forecast estimates. Panel A predicts annual equity premia. Necessarily, all *ex-post* λ combinations have positive ΔRMSE*—but almost all rolling λ combinations have negative ΔRMSE$^{\star r}$. The exceptions are **d/e** and **cayp** (with OOS knowledge). In some but not all specifications, this also applies to **dfy**, **all**, and **caya**. **d/e**, **dfy**, and **all**can immediately be excluded, because their optimal λ is negative. This leaves **caya**. Again, not reported, **caya** could not outperform over the most recent three decades. In the monthly rolling encompassing tests (not reported), only **svar** and **d/e** (in one specification) are positive, neither with a positive λ.

[12] For the first three observations, we presume perfect optimal foresight, resulting in the minimum ΔRMSE. This tilts the rolling statistic slightly in favor of superior performance. The results remain the same if we use reasonable variations.

Table 6
Encompassing tests
This table presents statistics on encompassing tests for excess stock return forecasts at various frequencies. Variables are explained in Section 1. All numbers are in percent per frequency corresponding to the panel. λ gives the ex-post weight on the conditional forecast for the optimal forecast that minimizes the MSE. ENC is the test statistic proposed by Clark and McCracken (2001) for a test of forecast encompassing. One-sided critical values of ENC statistic are obtained empirically from bootstrapped distributions, except for caya, cayp, and all models where they are obtained from Clark and McCracken (2001). Critical values for ms model are not calculated. cayp uses ex-post information. ΔRMSE* is the RMSE difference between the unconditional forecast and the optimal forecast for the same sample/forecast period. $\Delta RMSE^{*r}$ is the RMSE difference between the unconditional forecast and the optimal forecast for the same sample/forecast period using rolling estimates of λ. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

		Estimation:	All data					All Data				After 1927				
		OOS Forecast:	After 20 years					After 1965				After 1965				
		Data	$\overline{R}^2$	λ	ENC	ΔRMSE*	$\Delta RMSE^{*r}$	λ	ENC	ΔRMSE*	$\Delta RMSE^{*r}$	$\overline{R}^2$	λ	ENC	ΔRMSE*	$\Delta RMSE^{*r}$
d/p	Dividend Price Ratio	1872–2005	0.49	0.21	0.48	+0.0084	−0.2583	0.40	**0.87***	+0.0664	−0.4989	1.67	0.54	**2.19****	+0.2297	−0.3539
d/y	Dividend Yield	1872–2005	0.91	0.38	1.94	+0.0614	−0.5713	0.30	**1.24***	+0.0749	−0.5389	**2.71***	0.41	**3.24****	+0.2662	−0.2858
e/p	Earning price ratio	1872–2005	1.08	0.22	0.40	+0.0074	−0.2266	0.66	**1.21****	+0.1508	−0.4845	**3.20***	0.48	**2.51****	+0.2346	−0.4049
d/e	Dividend payout ratio	1872–2005	−0.75	−1.73	−1.46	+0.2135	+0.0960	−8.46	−0.45	+0.7545	+0.2858	−1.24	−4.57	−1.25	+1.2308	+0.7796
svar	Stock variance	1885–2005	−0.76	−0.42	−4.74	+0.2387	−0.6475	2.07	0.03	+0.0134	−0.5937	−1.32	−16.73	−0.18	+0.5906	+0.4490
b/m	Book to market	1921–2005	**3.20***	0.49	**4.16****	+0.2532	−0.0575	0.20	**1.27***	+0.0559	−0.7885	**4.14***	0.18	**1.67***	+0.0689	−0.4821
ntis	Net equity expansion	1927–2005	**8.15*****	0.31	1.46	+0.0619	−0.2708	0.31	**1.30***	+0.0805	−0.9310	**8.15*****	0.31	**1.30***	+0.0805	−0.9310
eqis	Pct equity issuing	1927–2005	**9.15*****	0.67	**4.45****	+0.3917	−0.0564	0.56	**3.12****	+0.3342	−0.7106	**9.15*****	0.56	**3.12****	+0.3342	−0.7106
tbl	Treasury-bill rate	1920–2005	0.34	0.39	**2.14***	+0.1031	−1.2425	0.41	**2.16****	+0.1790	−1.3058	0.15	0.33	**2.72****	+0.1863	−0.5619
lty	Long term yield	1919–2005	−0.63	0.29	**2.67***	+0.0971	−0.7012	0.28	**2.39****	+0.1447	−0.9358	−0.94	0.25	**2.39****	+0.1317	−0.5682
ltr	Long term return	1926–2005	0.99	0.31	**4.55****	+0.2077	−0.1412	0.24	**2.45****	+0.1300	−8.4290	0.92	0.25	**2.44****	+0.1348	−8.7284
tms	Term Spread	1920–2005	0.16	0.38	0.93	+0.0433	−1.0292	0.47	**1.07***	+0.0977	−0.8750	0.89	0.50	**1.95****	+0.1880	−0.5375
dfy	Default yield spread	1919–2005	−4.18	−2.62	−0.48	+0.1503	−0.9718	−10.65	−0.30	+0.6395	+0.4999	−1.31	−11.91	−0.24	+0.5677	+0.4496
dfr	Default Return Spread	1926–2005	0.40	0.44	0.87	+0.0501	−0.3698	0.47	0.78	+0.0710	−0.3808	0.32	0.48	0.74	+0.0692	−0.3877
infl	Inflation	1919–2005	−1.00	−2.46	−0.68	+0.2019	−0.4520	−1.48	−0.15	+0.0429	−15.1368	−0.99	−3.12	−0.86	+0.5541	−0.4697
i/k	Invstmnt capital ratio	1947–2005	**6.63****	0.53	**3.01****	+0.3330	−0.1950	0.53	**3.01****	+0.3330	−0.1950	**6.63****	0.53	**3.01****	+0.3330	−0.1950
cayp	Cnsmptn, wlth, incme	1945–2005	**15.72*****	1.34	**7.62*****	+1.7225	+0.3315	1.34	**7.62*****	+1.7225	+0.3315	**15.72*****	1.34	**7.62*****	+1.7225	+0.3315
all	Kitchen sink	1927–2005	**13.81****	0.13	4.86	+0.1607	+0.0160	−0.07	−1.26	+0.0342	−0.4666	**13.81****	−0.07	−1.26	+0.0342	−0.4666
caya	Cnsmptn, wlth, incme	1945–2005	-	0.45	**3.39****	+0.3117	−0.3185	0.45	**3.39****	+0.3117	−0.3185	-	0.45	**3.39****	+0.3117	−0.3185
ms	Model Selection	1927–2005	-	0.24	4.82	+0.1870	+0.0739	0.07	0.59	+0.0094	−1.1268	-	0.07	0.59	+0.0094	−1.1268

Table 6
panel B: Monthly Data

		OOS Forecast:	After 194701					After 196501			
		Data	$\overline{R}^2$	λ	ENC	ΔRMSE*	ΔRMSE*^{r}	λ	ENC	ΔRMSE*	ΔRMSE*^{r}
d/p	Dividend price ratio	192701–200512	0.15	0.53	**4.14****	+0.0065	−0.0134	0.53	**2.67****	+0.0063	−0.0109
d/y	Dividend yield	192701–200512	**0.25***	0.43	**6.53*****	+0.0083	−0.0115	0.45	**3.90****	+0.0078	−0.0084
e/p	Earning price ratio	192701–200512	**0.54****	0.35	**9.27*****	+0.0097	−0.0135	0.28	**3.08****	+0.0039	−0.0172
d/e	Dividend payout ratio	192701–200512	0.01	−0.02	−0.22	+0.0000	−0.0146	−1.12	−3.01	+0.0152	+0.0003
svar	Stock variance	192701–200512	−0.08	−12.30	−0.47	+0.0172	+0.0046	−12.93	−0.32	+0.0184	+0.0060
csp	Cross-sectional prem	193705–200212	**0.92*****	0.38	**6.21*****	+0.0093	−0.0138	0.82	**5.50*****	+0.0219	−0.0007
b/m	Book to market	192701–200512	**0.40****	0.18	**3.04****	+0.0016	−0.0416	0.07	0.89	+0.0003	−0.0260
ntis	Net equity expansion	192701–200512	**0.75*****	0.60	**4.28****	+0.0075	−0.0055	0.47	**2.77****	+0.0058	−0.0180
tbl	Treasury-bill rate	192701–200512	0.11	0.50	**5.47****	+0.0081	−0.0222	0.51	**4.86*****	+0.0110	−0.0218
lty	Long term yield	192701–200512	−0.01	0.35	**7.57*****	+0.0079	−0.0084	0.35	**5.47*****	+0.0086	−0.0161
ltr	Long term return	192701–200512	0.04	−0.15	−0.77	+0.0003	−4.0129	0.30	**1.02***	+0.0014	−0.0234
tms	Term spread	192701–200512	0.07	0.68	**2.51****	+0.0050	−0.0311	0.73	**2.37****	+0.0076	−0.0538
dfy	Default yield spread	192701–200512	−0.07	−1.04	−0.27	+0.0008	−0.0070	2.15	0.20	+0.0019	−0.0197
dfr	Default return spread	192701–200512	−0.02	−0.85	−0.72	+0.0018	−0.0134	−0.03	−0.01	+0.0000	−0.0221
infl	Inflation	192701–200512	−0.00	1.01	0.69	+0.0021	−0.0114	1.19	0.58	+0.0030	−0.0541
all	Kitchen sink	192701–200512	**1.98*****	0.05	4.39	+0.0008	−0.0150	0.14	**5.88****	+0.0040	−0.0366
ms	Model selection	192701–200512	–	0.09	1.51	+0.0004	−0.0232	0.14	1.39	+0.0009	−0.0245

In sum, "learned shrinking" does not improve any of our models to the point where we would expect them to outperform.

7. Other Literature

Our article is not the first to explore or to be critical of equity premium predictions. Many bits and pieces of evidence we report have surfaced elsewhere, and some authors working with the data may already know which models work, and when and why—but this is not easy to systematically determine for a reader of this literature. There is also a publication bias in favor of significant results—nonfindings are often deemed less interesting. Thus, the general literature tenet has remained that the empirical evidence and professional consensus is generally supportive of predictability. This is why we believe that it is important for us to review models in a comprehensive fashion—variable-wise, horizon-wise, and time-wise—and to bring all variables up-to-date. The updating is necessary to shed light on post-Oil Shock behavior and explain some otherwise startling disagreements in the literature.

There are many other articles that have critiqued predictive regressions. In the context of dividend ratios, see, for example, Goetzmann and Jorion (1993) and Ang and Bekaert (2003). A number of articles have also documented low IS power [e.g., see Goetzmann and Jorion (1993), Nelson and Kim (1993), and Valkanov (2003)). We must apologize to everyone whose article we omit to cite here—the literature is simply too voluminous to cover fully.

The articles that explore model instability and/or OOS tests have the closest kinship to our own. The possibility that the underlying model has changed (often through regime shifts) has also been explored in such articles as Heaton and Lucas (2000), Jagannathan, McGrattan and Scherbina (2000), Bansal, Tauchan and Zhou (2003), and Kim, Morley, and Nelson (2005), and Lettau and Nieuwerburgh (2005). Interestingly, Kim, Morley, and Nelson (2005) cannot find any structural univariate break post WW II. Bossaerts and Hillion (1999) suggest one particular kind of change in the underlying model—a disconnect between IS and OOS predictability because investors themselves are learning about the economy.

Again, many of the earlier OOS tests have focused on the dividend ratios.

- Fama and French (1988) interpret the OOS performance of dividend ratios to have been a success. Our article comes to the opposite conclusion primarily because we have access to a longer sample period.
- Bossaerts and Hillion (1999) interpret the OOS performance of the dividend yield (not dividend price ratio) to be a failure, too. However, they rely on a larger cross-section of 14 (correlated) countries and not

on a long OOS time period (1990–1995). Because this was a period when the dividend yield was known to have performed poorly, the findings were difficult to generalize.
- Ang and Bekaert (2003) similarly explore the dividend yield in a more rigorous structural model. They, too, find poor OOS predictability for the dividend yield.
- Goyal and Welch (2003) explore the OOS performance of the dividend ratios in greater detail on annual horizons. (Our current article has much overlap in perspective, but little overlap in implementation.)

Lettau and Ludvigson (2001) run rolling OOS regressions—but not in the same spirit as our article: the construction of their **cay** variable itself relies on *ex-post* coefficient knowledge. This thought experiment applies to a representative investor who knows the full-sample estimation coefficients for **cay**, but does not know the full-sample predictive coefficients. This is *not* the experiment our own article pursues. (Lettau and Ludvigson (2001) also do not explore their model's stability, or note its performance since 1975.) Some tests are hybrids between IS and OOS tests (as are our encompassing tests). For example, Fisher and Statman (2006) explore mechanical rules based on P/E and dividend-yield ratios, which are based on prespecified numerical cutoff values. None works robustly across countries.

Most of the above articles focus on a relatively small number of models. There are at least three studies in which the authors seek to explore more comprehensive sets of variables:

- Pesaran and Timmermann (1995) (and others) point out that our profession has snooped data (and methods) in search of models that seem to predict the equity premium in the same single U.S. or OECD data history. Their article considers model selection in great detail, exploring dividend yield, earnings–price ratios, interest rates, and money in $2^9 = 512$ model variations. Their data series is monthly, begins in 1954, and ends (by necessity) 12 years ago in 1992. They conclude that investors could have succeeded, especially in the volatile periods of the 1970s (i.e., the Oil Shock). But they do not entertain the historical equity premium mean as a NULL hypothesis, which makes it difficult to compare their results to our own. Our article shows that the Oil Shock experience generally is almost unique in making many predictive variables seem to outperform. Still, even including the 2-year Oil Shock period in the sample, the overall OOS performance of our ALTERNATIVE models is typically poor.
- Ferson, Sarkissian and Simin (2003) explore spurious regressions and data mining in the presence of serially correlated independent variables. They suggest increasing the critical t-value of the IS regression. The article concludes that "many of the regressions in the literature, based on individual predictor variables, may be spurious."

Torous and Valkanov (2000) disagree with Ferson, Sarkissian, and Simin. They find that a low signal–noise ratio of many predictive variables makes a spurious relation between returns and persistent predictive variables unlikely and, at the same time, would lead to no OOS forecasting power.

- An independent study, Rapach and Wohar (2006), is perhaps closest to our article. It is also fairly recent, fairly comprehensive, and explores OOS performance for a number of variables. We come to many similar conclusions. Their study ends in 1999, while our data end in 2005—a fairly dramatic five years. Moreover, our study focuses more on diagnosis of weaknesses, rather than just on detection.[13]

8. Conclusion

Findings: Our article systematically investigates the IS and OOS performance of (mostly) linear regressions that predict the equity premium with prominent variables from earlier academic research. Our analysis can be regarded as conservative because we do not even conduct a true OOS test—we select variables from previously published articles and include the very same data that were used to establish the models in the first place. We also ignore the question of how a researcher or investor would have known which among the many models we considered would ultimately have worked.

There is one model for which we feel judgment should be reserved (**eqis**), and some models that deserve more investigation on very-long term frequencies (5 years). None of the remaining models seems to have worked well. To draw this conclusion, our article relies not only on the printed tables in this final version, but on a much larger set of tables that explore combinations of modified data definitions, data frequencies, time periods, econometric specifications, etc).[14] Our findings are not driven by a few outlier years. Our findings do not disappear if we use different definitions and corrections for the time-series properties of the independent variable. Our findings do not arise because our tests have weak power (which would have manifested itself mostly in poor early predictions). Our findings hold up if we apply statistical corrections, data driven model selection, and encompassing tests.

[13] Another study by Guo (2006) finds that **svar** has OOS predictive power. However, Guo uses post WW II sample period and downweights the fourth quarter of 1987 in calculating stock variance. We check that this is why he can find significance where we find none. In the pre-WW2 period, there are many more quarters that have even higher stock variance than the fourth quarter of 1987. If we use a longer sample period, Guo's results also disappear regardless of whether we downweight the highest observation or not.

[14] The tables in this article have been distilled from a larger set of tables, which are available from our website—and on which we sometimes draw in our text description of results.

Instead, our view based on this evidence is now that most models seem unstable or even spurious. Our plots help diagnose when they performed well or poorly, both IS and OOS. They shine light on the two most interesting subperiods, the 1973–75 Oil Shock, and the most recent 30 years, 1975 till today. (And we strongly suggest that future articles proposing equity premium predictive models include similar plots.) If we exclude the Oil Shock, most models perform even worse—many were statistically significant in the past only because of the stellar model performance during these contiguous unusual years. One can only imagine whether our profession would have been equally comfortable rationalizing away these years "as unusual" if they had been the main negative and not the main positive influence.

As of the end of 2005, most models have lost statistical significance, both IS and OOS. OOS, most models not only fail to beat the unconditional benchmark (the prevailing mean) in a statistically or economically significant manner, but underperform it outright. If we focus on the most recent decades, that is, the period after 1975, we find that no model had superior performance OOS and few had acceptable performance IS. With 30 years of poor performance, believing in a model today would require strong priors that the model is well specified and that the underlying model has not changed.

Of course, even today, researchers can cherry-pick models—intentionally or unintentionally. Still, this does not seem to be an easy task. It is rare that a choice of sample start, data frequency, and method leads to robust superior statistical performance IS. Again, to ignore OOS tests even as a diagnostic, a researcher would have to have supreme confidence that the underlying model is stable. Despite extensive search, we were unsuccessful in identifying any models on annual or shorter frequency that systematically had both good IS and OOS performance, at least in the period from 1975 to 2005—although more search might eventually produce one. To place faith in a model, we would want to see genuine superior and stable IS and OOS performance in years after the model identification. Switching perspective from a researcher to an investor, we believe the evidence suggests that none of the academic models we reexamine warrants a strong investment endorsement today. By assuming that the equity premium was "like it always has been," an investor would have done just as well.

Directions: An academic researcher could explore more variables and/or more sophisticated models (e.g., through structural shifts or Kalman filters). Alternatively, one could predict disaggregated returns, for example, the returns on value stocks and the returns on growth stocks. The former could respond more strongly to dividends, while the latter could respond more strongly to book-to-market factors. However, such explorations aggravate the problems arising from (collective) specification search. Some of these models are bound to work both IS or OOS by pure chance. At

the very least, researchers should wait for more new OOS data to become available in order to accumulate faith in such new variables or more sophisticated models.

Having stated the obvious, there are promising directions. We are looking forward to accumulating more data. Lettau and Van Nieuwerburgh (2005) model structural change not on the basis of the forecasting regression, but on the basis of mean shifts in the dependent variables. This reduces (but does not eliminate) snooping bias. Another promising method relies on theory—an argument along the line of Cochrane's (2005) observation that the dividend yield must predict future returns eventually if it fails to predict dividend growth.[15]

Broader Implications: Our article is simple, but we believe its implications are not. The belief that the state variables that we have explored in our article can predict stock returns and/or equity premia is not only widely held, but the basis for two entire literatures: one literature on how these state variables predict the equity premium and one literature on how smart investors should use these state variables in better portfolio allocations. This is not to argue that an investor would not update his estimate of the equity premium as more equity premium realizations come in. Updating will necessarily induce time-varying opportunity sets [see Xia (2001) and Lewellen and Shanken (2002)). Instead, our article suggests only that the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power, both IS and OOS. We hope that the simplicity of our approach strengthens the credibility of our evidence.

Website Data Sources

Robert Shiller's Website: http://aida.econ.yale.edu/~shiller/data.htm.
NBER Macrohistory Data Base: http://www.nber.org/databases/macrohistory/contents/chapter13.html.
FRED: http://research.stlouisfed.org/fred2/categories/22.
Value-Line: http://www.valueline.com/pdf/valueline_2005.pdf.
Bureau of Labor Statistics Webpage: http://www.bls.gov/cpi/
Martin Lettau's Webpage: (cay), http://pages.stern.nyu.edu/~mlettau/.
William Schwert's Webpage: (svar), http://schwert.ssb.rochester.edu/.
Jeff Wurgler's Webpage: (eqis), http://pages.stern.nyu.edu/~jwurgler/

[15] We do not agree with all of Cochrane's (2005) conclusions. He has strong priors, placing full faith in a stationary specification of the underlying model—even though Goyal and Welch (2003) have documented dramatic increases in the autocorrelation of dividend growth. Therefore, he does not consider whether changes in the model over the last 30 years could lead one to the conclusion that dividend ratios do not predict *as of 2006*. He also draws a stark dichotomy between a NULL (no return prediction, but dividend growth prediction) and an ALTERNATIVE (no dividend growth prediction, but return prediction). He evaluates both hypotheses separately for dividend growth and return predictability. He then proceeds under unconditional confidence in the ALTERNATIVE to show that if dividend growth rates are truly unpredictable, then dividend ratios increase in significance to conventional levels. With residual doubts about the ALTERNATIVE, this conclusion could change.

References

Ang, A., and G. Bekaert, 2003. *Stock Return Predictability: Is it There?*, Working Article, Columbia University forthcoming in RFS.

Avramov, D. 2002. Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics*, 64(3):423–58.

Baker, M., and J. Wurgler, 2000. The equity share in new issues and aggregate stock returns. *Journal of Finance*, 55(5):2219–57.

Baker, M., R. Taliaferro, and J. Wurgler, 2004. Pseudo Market Timing and Predictive Regressions, NBER Working paper No. 10823.

Ball, R., 1978. Anomalies in Relationship Between Securities' Yields and Yield-Surrogates. *Journal of Financial Economics*, 6(2/3):103–26.

Bansal, R., G. Tauchen, and H. Zhou, 2004. Regime-shifts, risk premiums in the term structure, and the business cycle. *Journal of Business & Economic Statistics*, 22(4):396–409.

Bossaerts, P., and P. Hillion, 1999. Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies*, 12(2):405–28.

Boudoukh, J., M. Richardson, and R. F. Whitelaw, 2005. The Myth of Long-Horizon Predictability, NBER Working Paper No. 11841.

Boudoukh, J., R. Michaely, M. P. Richardson, and M. R. Roberts, 2007. On the importance of measuring payout yield: implications for empirical asset pricing. *Journal of Finance*, 62(2):877–915.

Breen, W., L. R. Glosten, and R. Jagannathan, 1989. Economic significance of predictable variations in stock index returns. *Journal of Finance*, 64(5):1177–89.

Brennan, M. J., and Y. Xia, 2004. Persistence, Predictability, and Portfolio Planning, Working Paper, UCLA and Wharton.

Butler, A. W., G. Grullon, and J. Weston, 2005. Can managers forecast aggregate market returns?. *Journal of Finance*, 60(2):963–86.

Campbell, J. Y., 1987. Stock returns and the Term Structure. *Journal of Financial Economics*, 18(2):373–99.

Campbell, J. Y., and R. J. Shiller, 1988a. Stock prices, earnings, and expected dividends. *Journal of Finance*, 43(3):661–76.

Campbell, J. Y., and R. J. Shiller, 1988b. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–227.

Campbell, J. Y., and R. J. Shiller, 1998. Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, 24(2):11–26.

Campbell, J. Y., and L. M. Viceira, 2002. *Strategic Asset Allocation: Portfolio Choice for Long-term Investors*, Oxford University Press, Oxford.

Campbell, J. Y., and T. Vuolteenaho, 2004. Inflation illusion and stock prices. *American Economic Review*, 94(2):19–23.

Campbell, J. Y., and S. B. Thompson, 2005. *Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?*, Working Paper, Harvard University Forthcoming in RFS.

Campbell, J. Y., and M. Yogo, 2006. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Clark, T. E., and M. W. McCracken, 2001. Tests of forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105:85–110.

Cochrane, J. H., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance*, 46(1):209–37.

Cochrane, J. H., 1997. Where is the market going? Uncertain facts and novel theories. *Federal Reserve Bank of Chicago - Economic Perspectives*, 21(6):3–37.

Cochrane, J. H., 2005. *The Dog That Did Not Bark: A Defense of Return Predictability*, Working Paper, University of Chicago Forthcoming in RFS.

Diebold, F. X., and R. S. Mariano, 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

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

Fama, E. F., 1981. Stock returns, real activity, inflation, and money. *American Economic Review*, 71(4):545–65.

Fama, E. F., and G. W. Schwert, 1977. Asset returns and inflation. *Journal of Financial Economics*, 5(2):115–46.

Fama, E. F., and K. R. French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F., and K. R. French, 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Ferson, W. E., S. Sarkissian, and T. T. S. Simin, 2003. Spurious regressions in financial economics. *Journal of Finance*, 58(4):1393–413.

Fisher, K. L., and M. Statman, 2006. Market timing at home and abroad, *Journal of investing*, 2006, 15(2):19–27.

Goetzmann, W. N., and P. Jorion, 1993. Testing the predictive power of dividend yields. *Journal of Finance*, 48(2):663–79.

Goyal, A., and I. Welch, 2003. Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–54.

Guo, H., 2006. On the out-of-sample predictability of stock market returns. *Journal of Business*, 79(2):645–70.

Harvey, D. I., S. J. Leybourne, and P. Newbold, 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–91.

Heaton, J., and D. J. Lucas, 2000. Stock prices and fundamentals, in B. S. Bernanke, and J. Rotemberg (ed.), *NBER Macroeconomics Annual 1999*, Vol. 14, MIT Press/NBER, Cambridge, MA, 213–42.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies*, 5(3):257–86.

Inoue, A., and L. Kilian, 2004. In-sample or out-of-sample tests of predictability: which one should we use?, *Econometric Reviews*, 23(4):371–402.

Jagannathan, R., E. R. McGrattan, and A. Scherbina, 2000. The declining U.S. equity premium. *Federal Reserve Bank of Minneapolis Quarterly Review*, 24(4):3–19.

Keim, D. B., and R. F. Stambaugh, 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17(2):357–90.

Keim, D., and A. Madhavan, 1997. Transaction costs and investment style: an inter-exchange analysis of institutional equity trades. *Journal of Financial Economics*, 46(3):265–92.

Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long–horizon regressions? *Journal of Applied Econometrics*, 14(5):491–510.

Kim, C.-J., J. C. Morley, and C. R. Nelson, 2005. The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2):181–91.

Kothari, S., and J. Shanken, 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. *Journal of Financial Economics*, 44(2):169–203.

Lamont, O., 1998. Earnings and expected returns. *Journal of Finance*, 53(5):1563–87.

Lamoureux, C. G., and G. Zhou, 1996. Temporary components of stock returns: what do the data tell us?. *Review of Financial Studies*, 9(4):1033–59.

Lettau, M., and S. Ludvigson, 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56(3):815–49.

Lettau, M., and S. Ludvigson, 2005. Expected returns and expected dividend growth. *Journal of Financial Economics*, 76(3):583–626.

Lettau, M., and S. Van Nieuwerburgh, 2005. Reconciling the Return Predictability Evidence, Working Paper, NYU.

Lewellen, J., 2004. Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–35.

Lewellen, J., and J. Shanken, 2002. Learning, asset-pricing tests, and market efficiency. *Journal of Finance*, 57(3):1113–45.

Lintner, J., 1975. Inflation and security returns. *Journal of Finance*, 30(2):259–80.

Mark, N. C., 1995. Exchange rates and fundamentals: evidence on long–horizon predictability. *American Economic Review*, 85(1):201–18.

McCracken, M. W., 2004. *Asymptotics for Out-of-sample Tests of Causality*, Working Paper, University of Missouri-Columbia.

Menzly, L., T. Santos, and P. Veronesi, 2004. Understanding predictability. *Journal of Political Economy*, 112(1):1–47.

Nelson, C. R., and M. J. Kim, 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance*, 48(2):641–61.

Pesaran, H. M., and A. Timmermann, 1995. Predictability of stock returns: robustness and economic significance. *Journal of Finance*, 50(4):1201–28.

Polk, C., S. Thompson, and T. Vuolteenaho, 2006. Cross-sectional forecasts of the equity premium. *Journal of Financial Economics*, 81(1):101–41.

Pontiff, J., and L. D. Schall, 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics*, 49(2):141–60.

Rapach, D. E., and M. E. Wohar, 2006. In-Sample vs. Out-of-Sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, 13(2):231–47.

Rissanen, J., 1986. Order estimation by accumulated prediction errors. *Journal of Applied Probability*, 23A:55–61.

Rozeff, M. S., 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management*, 11(1):68–75.

Shiller, R. J., 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 2:457–98.

Stambaugh, R. F., 1999. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Torous, W., and R. Valkanov, 2000. Boundaries of Predictability: Noisy Predictive Regressions, Working Paper, UCLA.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, 68(2):201–32.

Xia, Y., 2001. Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation. *Journal of Finance*, 56(1):205–46.