A Comprehensive Look at The Empirical Performance of Equity Premium Prediction*

Amit Goyal
Goizueta Business School
Emory University†

Ivo Welch
Brown University
Department of Economics
NBER‡

September 19, 2005

Abstract

Economists have suggested a whole range of variables that predict the equity premium: dividend price ratios, dividend yields, earnings-price ratios, dividend payout ratios, corporate or net issuing ratios, book-market ratios, beta premia, interest rates (in various guises), and consumption-based macroeconomic ratios (cay). Our paper comprehensively reexamines the performance of these variables, both in-sample and out-of-sample. We find that most variables would not have helped an investor outpredict the historical equity premium mean. Most would have outright hurt. None deserves an unqualified endorsement.

JEL Classification: G12, G14.

*Thanks to Malcolm Baker, Ray Ball, John Campbell, Francis Diebold, Ravi Jagannathan, Owen Lamont, Sydney Ludvigson, Rajnish Mehra, Michael Roberts, 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 iterating drafts and exchanging perspectives with (or against) our earlier drafts—this has allowed us to significantly improve.

†http://www.goizueta.emory.edu/agoyal. E-mail: amit_goyal@bus.emory.edu.

‡http://welch.econ.brown.edu. E-mail: ivo_welch@brown.edu.

1 Introduction

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

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

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

Dividend-price ratio and dividend yield: Ball (1978), Rozeff (1984), Shiller (1984), Campbell (1987), Campbell and Shiller (1988a), Campbell and Shiller (1988b), Fama and French (1988), Hodrick (1992), Campbell and Viceira (2002), Campbell and Yogo (2003), Lewellen (2004), and Menzly, Santos, and Veronesi (2004). Cochrane (1997) surveys the dividend ratio prediction literature.

Earnings price ratio and dividend-earnings (payout) ratio: Campbell and Shiller (1988a), Campbell and Shiller (1998), and Lamont (1998), originally motivated by Graham and Dodd.

Interest and inflation rates: The short term interest rate: Campbell (1987) and Hodrick (1992). The term spread and the default spread: Avramov (2002), Campbell (1987), Fama and French (1989), and Keim and Stambaugh (1986). The inflation rate: Campbell and Vuolteenaho (2004), Fama (1981), Fama and Schwert (1977), and Lintner (1975). Some papers explore multiple interest rate related variables, as well as dividend related variables (e.g., Ang and Bekaert (2003)).

Book-to-market ratio: Kothari and Shanken (1997) and Pontiff and Schall (1998).

Consumption, wealth, and income ratio (CAY): Lettau and Ludvigson (2001).

Aggregate net issuing activity: Baker and Wurgler (2000) and Boudoukh, Michaely, Richardson, and Roberts (2003).

The literature is difficult to absorb. Different papers use different techniques, variables, and time periods. Many papers were written years ago, and thus did not have access to more recent data. Some papers contradict the findings of others. Still, most readers are left with the impression that "prediction works"—though it is less clear what exactly 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 working papers, forthcoming papers, and recently published papers, which 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 state variables—and prominently the just-mentioned variables.

The goal of our own paper is to comprehensively reexamine the empirical evidence as of 2004, examining each variable using the same method, time-period, and estimation frequency, with both in-sample (IS) and out-of-sample (OOS) tests. In brief, our results suggest broadly that none of the equity prediction models we examine are robust. None has worked during the most recent three decades.

Section 2 describes our data. Section 3 describes the tests we are performing. Section 4 explores our base case—predicting equity premia annually using ols forecasts. Section 5 predicts equity premia on five-year horizons. Section 6 predicts monthly equity premia, with special emphasis on the suggestions in Campbell and Thompson (2005). Section 7 tries earnings and dividend ratios with longer memory as independent variables and explores the Stambaugh (1999), Lewellen (2004), and Campbell and Yogo (2003) corrections. Section 8 puts "encompassing" model forecasts to the test. Section 9 reviews earlier literature. Section 10 summarizes, and speculates why these models performed so poorly.

2 Data Sources and Data Construction

We first describe our data sources and data construction. The dependent variable is always the equity premium, i.e., 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 2004 from CRSP's 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.

- Risk-free Rate: The risk-free rate for the period 1920 to 2004 is the T-bill rate. Because there was no risk-free short-term debt prior to the 1920's, we had to estimate it. We obtained commercial paper rates for New York City from NBER's Macrohistory data base. These are available for the period 1871 to 1970. We estimated a regression for the period 1920 to 1971, which yielded

$$\text{T-bill Rate} \;=\; -0.004 \;+\; 0.886 \cdot \text{Commercial Paper Rate}\;, \tag{2}$$

  with an R^2 of 95.7%. Therefore, we instrumented the risk-free rate for the period 1871 to 1919 with the predicted regression equation. The correlation for the period

1920 to 1971 between the *equity premium* computed using the T-bill rate and that computed using the predicted commercial paper rate is 99.8%.

The equity premium had a mean of 4.77%, median of 6.51%, and standard deviation of 17.88% over the entire sample period of 1872 to 2004. The equity premium is 5.99% (standard deviation of 19.31%) from 1927–2004, 6.35% (standard deviation of 15.89%) from 1947–2004, and 3.89% (standard deviation of 15.92%) from 1965–2004.

Our first set of independent variables relate primarily to characteristics of stocks:

- Dividends: Dividends are twelve-month moving sums of dividends paid on the S&P 500 index. They are from Robert Shiller's website for the period 1871 to 1970. Dividends from 1971 to 2004 are from 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.

- Earnings: Earnings are twelve-month moving sums of earnings on the S&P 500 index. These are from Robert Shiller's website for the period 1871 to June 2003. Earnings from June 2003 to December 2004 are our own estimates based on interpolation of quarterly earnings provided by S&P Corporation. The **Earnings Price Ratio** (**e/p**) is the difference between the log of earnings and the log of prices. (An occasional variation is $\mathbf{e}^{10}/\mathbf{p}$, which is a 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.

- Stock Variance (**svar**): Stock Variance is computed as sum of squared daily returns on S&P 500. Daily returns for 1871 to 1926 are from Bill Schwert. Daily returns from 1926 to 2004 are from CRSP.

- Cross-Sectional Premium (**csp**): The cross-sectional beta premium measures the relative valuations of high- and low-beta stocks. This is the variable proposed in Polk, Thompson, and Vuolteenaho (2006), which we obtained directly from Sam Thompson. This variable is available from May 1937 to December 2002.

- Book Value: Book values from 1920 to 2004 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 of March to December, this is computed by dividing book value at the end of previous year by the price at the end of the current month. For the months of January to February, this is computed by dividing book value at the end of 2 years ago by the price at the end of the current month.

- Net Issuing Activity: The dollar amount of net issuing activity (IPOs, SEOs, stock repurchases, less dividends) for NYSE listed stocks is computed from CRSP data via the following equation:

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

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 2004. **Net Equity Expansion** (**ntis**): is the ratio of twelve-month moving sums of net issues by NYSE listed stocks divided by the total market capitalization of NYSE stocks. **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), which we obtained directly from the authors, except for 2004 which we added ourselves.

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

- T-bills (**tbl**): T-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 from NBER's Macrohistory data base. T-bill rates from 1934 to 2004 are the *3-Month Treasury Bill: Secondary Market Rate* from the economic research database at Federal Reserve Bank at St. Louis (FRED).

- Long Term Yield (**lty**): Long-term government bond yields for the period 1919 to 1925 is the *U.S. Yield On Long-Term United States Bonds* series from NBER's Macrohistory database. Yields from 1926 to 2004 are from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. Long Term Rate of Return (**ltr**): Long-term government bond returns for the period 1926 to 2004 are from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. The **Term Spread** (**tms**) is the difference between the long term yield on government bonds and the T-bill.

- Corporate Bond Returns: Long-term corporate bond returns for the period 1926 to 2004 are from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. Corporate Bond Yields: Yields on AAA- and BAA-rated bonds for the period 1919 to 2004 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 the return on long-term corporate bonds and returns on the long-term government bonds.

- Inflation (**infl**): Inflation is the *Consumer Price Index (All Urban Consumers)* for the period 1919 to 2004 from the Bureau of Labor Statistics. Because inflation information is released only in the following month, in our monthly regressions, we inserted one month of waiting before use.

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 themselves rolling in their independent variable construction (**cay** and **ms**).

[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.

- A "Kitchen Sink" Regression, named "**all**" 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**) is suggested in Lettau and Ludvigson (2001). Data for its construction is available from Martin Lettau's website at quarterly frequency from the second quarter of 1952 to the fourth quarter of 2004, and for annual frequency from 1948 to 2001. Lettau-Ludvigson estimate the following equation:

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

  where c is the aggregate consumption, a is the aggregate wealth, and y is the aggregate income. The estimates of the above equation provide **cay** $\equiv \widehat{\text{cay}}_t = c_t - \hat{\beta}_a \cdot a_t - \hat{\beta}_y \cdot y_t, \quad t = 1, \ldots, T$. 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).)

  Because the Lettau-Ludvigson measure of **cay** is constructed using look-ahead (*in-sample* regression coefficients), we created an equivalent measure that 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

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

  where the superscript on betas indicates that these are rolling estimates. This measure is called **caya** ("ante") to distinguish it from the traditional variable **cayp** constructed with look-ahead bias ("post").

- A model selection approach, named "**ms**." If there are K variables, we consider 2^K models essentially consisting of all possible combinations of variables. 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 amongst competing models in each time period t. This is also similar in spirit to the use of more conventional criteria (like R^2) in Pesaran and Timmerman (1995) (who do not entertain our NULL hypothesis). This model also shares a certain flavor with our encompassing tests (Section 8), in which we seek to find an optimal rolling combination between each model and an unconditional historical equity premium average.

The latter two models change every period, which renders an in-sample regression problematic. This is also why we did not include **caya** in the kitchen sink specification.

3 Empirical Procedure

All regressions are estimated using OLS. The in-sample significance of a regression is determined using the F-statistic, critical values of which are estimated using the bootstrap procedure described below. 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

$$\begin{aligned} \Delta\text{RMSE} &= \sqrt{\text{MSE}_{\mathsf{N}}} - \sqrt{\text{MSE}_{\mathsf{A}}} \quad , \\ \text{MSE-F} &= (T - h + 1) \cdot \left(\frac{\text{MSE}_{\mathsf{N}} - \text{MSE}_{\mathsf{A}}}{\text{MSE}_{\mathsf{A}}} \right) \quad , \end{aligned} \tag{6}$$

where h is the overlap degree ($h = 1$ for no overlap). MSE-F is the F-statistic by McCracken (2004). It tests for equal (R)MSE of the unconditional forecast and the conditional forecast (i.e., Δ(R)MSE $= 0$).[2] For our encompassing tests in Section 8, we compute

$$\text{ENC} = \frac{T - h + 1}{T} \cdot \frac{\sum_{t=1}^{T} \left(e_{\mathsf{N}t}^2 - e_{\mathsf{N}t} \cdot e_{\mathsf{A}t} \right)}{\text{MSE}_{\mathsf{A}}} \quad , \tag{7}$$

which is the statistic proposed by Clark and McCracken (2001) for an encompassing forecast test. They also show that these statistics follow non-standard distributions when testing nested models, because the asymptotic difference in squared forecast errors is exactly 0 with 0 variance under the NULL, which renders 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 five-year *overlapping* observations (for which asymptotic critical values are not available), we obtain critical values from the bootstrap procedure described below (critical values for **caya** and **all** models are not calculated using bootstrap, 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]

[2] Our earlier drafts also entertained another performance metric, the mean absolute error difference ΔMAE. The results were very similar. These drafts also described another standard error, MSE-T $= \sqrt{T + 1 - 2 \cdot h + h \cdot (h-1)/T} \cdot \left[(\overline{d}) / (\widehat{se}(\overline{d})) \right]$, where $d_t = e_{\mathsf{N}t} - e_{\mathsf{A}t}$, and $\overline{d} = T^{-1} \cdot \sum_t^T d_t = \text{MSE}_{\mathsf{N}} - \text{MSE}_{\mathsf{A}}$ 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 Harvey, Leybourne, and Newbold (1997). (We still use the latter as bounds in our plots, because we know the full distribution.) Again, the results were very similar. We chose to use the MSE-F in this paper because Clark and McCracken (2001) find that MSE-F has higher power than MSE-T.

[3] If the regression coefficient β is small (so that explanatory power is low or the in-sample 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.*

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

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

The bootstrap for calculating power assumes the data generating process is

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

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

We explore three time period specifications: the first begins OOS forecasts twenty years after data are available; the second begins OOS forecast in 1965; the third ignores all data prior to 1927 even in the estimation. If a variable does not have complete data, some of these time-specifications can overlap. Using three different periods reflect different tradeoffs between the desire to obtain statistical power, and the desire to obtain results that remain relevant today. The point estimates come from simple OLS, although, as noted earlier, their statistical significance is bootstrapped.

Table 1: Annual Performance

Table 1 explores predictive performance on annual forecasting horizons. Our tables strive to follow a common format. Panel A shows all models that have no in-sample significance at the *90% significance level*. Lack of IS significance obviates the need for OOS tests, although we do include the related ΔRMSE OOS information for comparability. Panel B describes in more detail those models that are statistically significant IS or for which there is no IS analog (**cayp** and **ms**).

4.1 Annual Prediction — Models In-Sample Insignificant

Panel A of Table 1 shows that most models in the literature have no IS explanatory power at the end of 2004, even at 90% significance level. In the full-data IS regression, even the two highest adjusted $\overline{R}^2$ models, **e/p** and **ltr**, have two-sided p-values of only 13% and 19%, respectively. If a model has no IS performance, its OOS performance is of course not too interesting—no investor would rely on such a model today. Nevertheless, the table shows the OOS performance of these models, and unsurprisingly none had superior performance.

Figure 1 graphs the IS and OOS performance of our 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 meaningful but the time-series pattern helps us diagnose the years of good or bad performance. Indeed, the OOS plots are sign-identical with the ΔRMSE statistic in our tables. (The rolling ΔRMSE statistic itself is too noisy to plot especially early on.) In some of our following discussions related to our plots, we refer to either number as OOS peformance. Again, the sign of the ΔRMSE reported in the first specification of Table 1 and the sign of the final observation in the figure are the same.

The standard error of the all observations is based on translating MSE-T statistic into symmetric 95% confidence intervals based on the McCracken (2004) critical values. Note that the test (standard errors) themselves are not asymptotically diminishing.

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 zero line). Indeed, a horizontal line and the right-side scale indicates the equivalent zero-point for the second time period specification, in which we begin forecasts in 1965. (The plots have also shifted the IS errors, so that the IS line begins at zero on the date of our first OOS prediction.) The Oil Shock NBER recession of 1973 to 1975 is marked by a vertical (red) bar in the figures.[4] Note also that the nadir and zenith—the years in which a model would have appeared worst and best, respectively—are invariant to the starting year (zero level) for OOS prediction, provided they remain in the sample.

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

1. both significant IS and OOS performance;
2. an irregular but upward drift;
3. an upward drift not just in one short or unusual sample period—say just the two years around the 1974 Oil Shock; and
4. 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 alternative explanation that the underlying model has changed.

[4] The actual recession period was from November 1973 to March 1975. We consider both 1974 and 1975 as years of Oil Shock recession in annual prediction.

Price Ratios (d/p, d/y, e/p). Figure 1 shows that the IS regressions for the three price ratios indicate reasonably steady performance. Table 1 Panel A shows that all three are not statistically significant IS at our 90% significance level—but the plots show why. The three ratios were significant in 1990, but are no longer so, not because the ΔRMSE has not recovered, but because the 1995–2002 period has added considerable volatility and thereby raised the standard errors. That is, Figure 1 shows that the IS models have recovered almost all of their predictive losses during the 1998-2000 bubble period when it comes to point estimates.[5]

However, the OOS regression performance of the price ratios is systematically worse than their IS performance. Figure 1 helps to diagnose their performances. **d/p** started off with bad performance from 1905 to WW-II, then had good performance beginning around WW-II, which ended in a plateau between 1975 and 1995. The most recent 30 years are interesting. Contrary to common perception, **d/p** did not perform poorly throughout, but only in the bubble period, 1995-2000, where its OOS performance dropped 0.0796, although it has regained 0.0183 by 2004. An important caveat must be that **d/p**'s good performance was entirely based on its predictive performance around the Oil Shock of 1973-1975. From 1973 to 1975, **d/p** gained 0.0477.

As noted earlier, these plots make it relatively easy to see what kind of sample periods would indicate in a regression whether a variable works or fails. 1936 and 1965 were two nadirs for **d/p**; 1984 was its zenith. Therefore, samples beginning around 1937 (or 1966) and ending in 1984 would suggest high statistical significance. As it turns out, 1965 was our starting year in the second specification, which is marked as the "Spec B Zero Val" line in the graph. **d/p**'s nadir-to-zenith performance was 0.0845. Comparing this to the aforementioned 0.0477 gain in the three consecutive years around the Oil Shock for perspective, it becomes obvious that more than half of **d/p**'s best performance was due to the Oil Shock. Also, the plot nicely shows that the most recent 30 years after the Oil Shock have not been kind to the **d/p** model:

d/p	All	Nadir–Zenith 1937–1984		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.0462	0.0845	0.0368	–0.0667
ΔRMSE	–0.11%	0.51%	0.25%	–0.79%

This further shows that a full 0.0477 of the best 0.0845 performance was due to the Oil Shock experience (1973–1975). Without it, even the maximum ΔRMSE would have been only half as high.

[5] The figures also shows that if we end the predictions in the mid-nineties, **d/p** would join the set of significant models. Indeed, an appendix table in our earlier draft showed that *only* **d/p**, **d/y**, and **e/p** join the set of variables with statistical significance IS (but not OOS) if we end the data in 1990. We will not dwell on these results and the dividend ratios, because Goyal and Welch (2003) focused on these. Moreover, earlier versions of our paper found qualitatively similar results for almost all variables if we end our data sample in 2002 or 2003.

Surprisingly, although the IS patterns of **d/p** and **d/y** look alike, **d/y**'s OOS pattern looks quite different. It is much more cyclical. It performed poorly until the Great Depression, well during the Great Depression, poorly in the New Deal, great from WW-II until about 1958, very poorly until the Oil Shock, great during the Oil Shock, and poorly all the way until the collapse of the bubble. Its best sample period would have been 1925 to 1957 for an 0.1372 gain. Extending this to the end of 1975 would leave a healthy 0.0926. For **d/y**, the Oil Shock was a good 0.0584, but it was not its only remarkable sample period. And, again, the most recent 30 years after the Oil Shock have not been kind to this variable:

d/y	All	Nadir–Zenith 1925–1957		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.0410	0.1372	NA	–0.0989
ΔRMSE	–0.10%	0.90%	NA	–1.15%

e/p's OOS pattern again looks different. It performed poorly until WW-II, and well from its nadir in 1942 to its zenith in 1976, gaining 0.0696. Extending the "optimal" sample end from 1976 to 1994 increase the OOS performance to 0.0849, which is significant. The plot shows that in order to find **e/p** to perform well, one must begin one's sample after the two disastrous forecast years 1917–1918 and prior to its good forecast period right after WW-II (around 1946). (Indeed, the periods 1942-1976, 1942-1994, and 1942-2004 have all statistically significant OOS performance at the 95% confidence level.) And, again, the most recent 30 years after the Oil Shock have not been kind to this variable:

e/p	All	Nadir–Zenith 1943–1976		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.0374	0.0696	0.0495	–0.0167
ΔRMSE	–0.09%	0.59%	0.52%	–0.20%

Other Variables The other plots in Figure 1 show that **d/e**, **ntis**, **dfy**, and **infl** never had significantly positive OOS periods, and that **svar** had a huge drop in OOS performance from 1930 to 1933. The other variables from Table 1 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 the 1990s that invalidated most models, with the exception of the aforementioned price ratio models.

4.2 Annual Prediction — Models In-Sample Significant

Panel B of Table 1 explores the performance of models that were either IS significant (**b/m**, **eqis**, and **all**), or for which there is no IS analog (**caya**, **ms**). This is done in more detail than in Panel A of the same Table. Specifically, we include statistics of IS ΔRMSE. This is done for the entire IS period and for the OOS period. This helps us judge the IS stability of the regression by comparing IS to OOS performance.

The book-market ratio (b/m) is statistically significant IS at the 6% level, regardless of whether we begin the regression in 1921 or 1927. (It does have positive IS performance in the forecast period 1941 to 2004, but not in the forecast period 1965 to 2004.)

Figure 2 shows why this is the case—and why **b/m** was deemed a successful predictor in past papers. It performs especially well if we can include the 1921–1941 period, which is only possible in the IS regressions. Moreover, it had reliable superior performance all the way from the Great Depression through the Oil Shock—and reliable inferior performance since then. OOS, its nadir was just at the outset (1941), its zenith was in 1976, for a gain of 0.1646. From 1976 to 2004, it lost 0.1513. If we begin forecasting in 1965, as in Table 1, **b/m**'s predictive underperformance is economically meaningful, with a ΔRMSE of –82.75 bp/year.

b/m	All	Nadir–Zenith 1942–1976		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.0079	0.1646	0.1654	–0.1513
ΔRMSE	–0.04%	1.43%	1.87%	–1.73%

As with our other models, its lack of OOS significance is not just a matter of low test power. For example, Table 1 Panel B shows that in the OOS prediction beginning in 1941, the OOS statistic came out *statistically significantly* positive in 68% of our (stable-model) simulations in which the IS regression was significant. In reality, we do not only not see statistical significance, we do not even see positive performance. Not reported in the table, positive performance (significant or insignificant) occurred in 81% of our simulations.

eqis is the clear standout among our models, with statistical significance both IS and OOS—even though our OOS model necessarily must exclude its IS stellar first fifteen years. Figure 2 shows that the IS and OOS predictions closely overlap—this model is therefore quite stable. Overall, the IS regression had a ΔRMSE of 1.06%/year, above the 0.43% IS ΔRMSE in the OOS period. Still, the loss due to model instability (from 0.43% IS to 0.38% OOS; and from 0.46% IS to 0.23% OOS) is very small, which allows **eqis**

to have statistically significant OOS performance. In sum, the rolling model would have outpredicted the prevailing historical mean statistically significantly and by between 23 and 37 bp/year.

Figure 2 has some surprises, though:

- Bad Years: **eqis** performed very poorly from 1937–1942, which was just the five years before the beginning of our OOS period.
- Neutral Years: **eqis** performed neither especially good nor bad after 1974—which gives it a big advantage relative to many other variables, which deteriorate badly. Like most of our other variables, it failed to predict the 2001 downturn, but did recover with the 2002 stock market.
- Good Years: **eqis** predicts exceptionally well in two periods: the Great Depression of 1929–1936 (i.e., long before our OOS period), and the Oil Shock of 1973 to 1975.

Repeating our earlier summary,

eqis	All	Nadir–Zenith 1949–1981		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.0690	0.1479	0.0405	–0.0531
ΔRMSE	0.12%	1.32%	0.45%	–0.63%

shows how $0.1479 - 0.0405 \approx 0.1074$ out of a maximum 0.1479 performance was due to the Oil Shock—a number which is larger than the total ΔSSE performance of 0.0690 for **eqis**. In any case, **eqis**has underperformed over the last three decades.

Our plot can also help explain dueling perspectives about **eqis** between Butler, Grullon, and Weston (2004) and Baker, Taliaferro, and Wurgler (2004). One part of their disagreement is whether **eqis**' performance is just random underperformance in sampled observations. Of course, some good years are expected to occur in any regression, but **eqis**' superior performance may not have been so random, because it [a] occurred in consecutive years, and [b] in response to the Oil Shock events that are often considered to have been exogenous, unforecastable, and unusual. Butler, Grullon, and Weston (2004) also end their data in 2002; while Baker, Taliaferro, and Wurgler (2004) refer to our earlier draft and Rapach and Wohar (2004), 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 papers have good points. We agree with Butler, Grullon, and Weston (2004) 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 collapse). Our conclusion is that while different papers can disagree about **eqis**, they should agree that it has not performed well over the most recent three decades.

all Panel B and Figure 2 show that the kitchen sink regression has high IS significance, but such inferior OOS performance that it dwarves all individual models combined.

cay Table 1 Panel B and Figure 2 also describe two models which have no in-sample analogs: **cay** and **ms** (the independent variables themselves require reestimation every period—the models change.) When we use the Lettau and Ludvigson (2001) proxy construction which takes advantage of ex-post information in the construction of **cay**, **cayp** has superior performance. This confirms the Lettau and Ludvigson experiment, in which their representative agent has knowledge not only of model parameters, but also of future consumption, wealth, and income data. Such an agent could have outpredicted the benchmark by 2.24% per annum:

cayp	All	Nadir–Zenith 1969–1997		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.2293	0.2622	0.0661	–0.0123
ΔRMSE	2.24%	3.09%	1.06%	–0.18%

Note also that 0.1961 of **cayp**'s 0.2622 maximum ΔSSE performance was due to the Oil Shock. However, if we use only ex-ante information in the construction of the **cay** proxy, **caya** cannot provide superior prediction.

caya	All	Nadir–Zenith 1969–1976		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.0553	0.1293	0.0162	–0.1800
ΔRMSE	–0.50%	3.72%	1.23%	–2.41%

And, regardless of whether one uses ex-post information, any superior performance of **cay** has ended in the mid 1970s. Even **cayp** had no really good years or good performance since then.

ms Finally, model selection **ms** fails with a similar pattern—good performance until 1976, dismal performance since.

ms	All	Nadir–Zenith 1949–1976		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	–0.1183	0.1939	0.1637	–0.2104
ΔRMSE	–0.63%	2.04%	2.34%	–2.36%

Conclusion In our sample period, 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 heavily on the mid-1970s Oil Shock for their apparent forecasting ability, often both IS and OOS. We caution against overreading or underreading this evidence. The fact that the important years are consecutive influential observations makes it unlikely that such observed behavior were "normal" and independent draws. At the same time, however, we do not know how we [the NULL or the ALTERNATIVE] could have known these special multi-year periods ahead of time, so predicting during these periods should not be easily discounted. More importantly and more unambiguously, no model seems to have performed well since—that is, over the last three decades.

On an annual prediction basis, there is no single variable that meets all of our suggested investment criteria:

- significant IS performance;
- positive, and preferably significant, OOS performance;
- reliable and reasonably steady predictive performance—not just based on the two Oil Shock years;
- positive, and preferably significant, OOS performance over the most recent three decades.

Most models fail on all four criteria.

5 Five-Yearly Frequencies

Some models may predict long-term returns better than short-term returns. Unfortunately, we do not have many years to thoroughly explore 5-year predictions, and therefore will only briefly touch on them. Table 2 repeats Table 1 with overlapping 5-year returns. As before, we bootstrap all critical significance levels. This is especially important here, because we now work with overlapping observations which leads to correlation in residuals.

For the most part, our annual inference carries over to the five-yearly frequencies. Panel A of Table 2 shows that most models remain in-sample statistically insignificant, excluding them from investment consideration in 2005. Panel B shows that **d/p** and **eqis** are the two IS statistically significant models, *provided* that we start estimation as soon as we have data; they are joined by **d/y** and **tms** if we start estimation in 1927. **d/p** and **d/y** have many of their good years prior to the beginning of the OOS period (the overall ΔRMSE is much higher than the IS for OOS ΔRMSE), and cannot significantly outperform the prevailing mean in OOS test. **eqis** now joins them in this failure. None of these three variables is a good 5-year predictor. Both the model selection and the kitchen sink again have dismal OOS performance.

There are two models/variables which seem to have positive OOS performance: **tms** and **caya**. The **tms** IS model even had its better years in the OOS period, with a ΔRMSE of 4.26% instead of the lower full IS ΔRMSE of 2.66%. This is enough to permit a superior OOS performance of 2.56% per five-years—a meaningful difference. An unreported plot shows that it performed well from 1968–1979, poorly from 1979–1986, and then well again from 1986–2004. Our concern with **tms** is that its performance is positive *only* if forecast begins in 1965, and negative if it begins in 1940—a small difference with a big effect.

The **caya** model has no-insample equivalent, but shows superior OOS performance, 2.06% per five-year period. On inspection, literally all its performance occurred in the 1997–1999 bubble period. For other variables with similar bubble gains, this gain is then dissipated from 2000–2002 when the market fell again. (Unfortunately, we do not have data to forecast 2002–2004 for **caya**.)

6 Monthly Prediction

Table 3 predicts monthly equity premia with variables available on a monthly basis. Because we require total returns from CRSP, we can only use data from 1927 onwards.

Although Table 3 Panel A shows that many of our models still remain statistically insignificant, the number of variables with IS statistical significance is considerably larger in the monthly than in the annual regressions. However, our conclusions are similarly pessimistic. Six out of the seven models that are either IS statistically significant or without IS analog (**ms**) underperform the prevailing mean OOS. The two models that outperform the prevailing mean are **ntis** if we begin our prediction in 1947, and **csp** if we begin our prediction in 1965. If we reverse the starting periods, the two models not only fail to outperform statistically significantly, they underperform. In sum, Table 3 does not point to a robust predictor of monthly equity premia.

6.1 Campbell-Thompson

We describe our further monthly evidence in the context of our discussion of Campbell and Thompson (2005) (CT), who offer a critique of our earlier drafts. The goal of this section is to reconcile their evidence and perspective with our own.

6.1.1 The Three Suggestions

CT highlight three important innovations, two of which can be seen as bringing "healthy distrust in the models" (our paper's point) into the models themselves.

1. They argue that a reasonable investor would not have used the models to forecast a negative equity premium. Therefore, they truncate such predictions at zero. (This can be applied to either IS and OOS analysis.) Given the high equity premium realizations in our sample period—and especially in the later half—this constraint naturally reduces poor predictions.

2. They argue that a reasonable investor would not have used a model that has a coefficient with an incorrect sign. Therefore, they truncate such coefficients at zero. (For some variables, such as the dividend ratios, this is easy. For other variables, however, it is not clear what the appropriate sign of the coefficient would be. The correct signs on coefficients in a multivariate regression are also not obvious. In any case, this restriction turns out not to have been important, as our table shows.)

These suggestions transform formerly linear models into non-linear models, which are generally not the subject of our paper.

Panel A of Table 4 follows the two CT suggestions, and adds some more diagnostics. The effect can be seen by comparing the plain $\Delta\text{RMSE}_{\text{PN}}$ and CT $\Delta\text{RMSE}_{\text{CT}}$ columns. The two CT restrictions do generally improve the OOS performance of the models (and therefore also their economic importance). However, among our IS significant models in Table 3, the two CT modifications change the OOS inference for only two models: **d/y** and **csp**. As we show in Figure 3 below, this occurs for **d/y** because it is now truncated to the unconditional model over the most recent decade, and therefore not of use to an investor in 2004. This is not unusual for **d/y**—52.1% of its predictions were truncated to zero. For **csp**, similarly, 50.2% of all months are not linearly predicted by **csp**, but truncated to zero. The remaining models remain qualitatively similar. That is, if they were significant or insignificant in the plain version, they remain so after the CT truncations. In sum, the two CT suggestions (and especially the first) can be recommended, but they cannot explain the difference in perspective.

The third CT suggestion pertains to the interpretation of magnitudes:

3. They propose using a certainty equivalence (CEV) measure to evaluate the out-of-sample predictive gains to a log-utility investor. (Brennan and Xia (2004) make a

similar argument.)

The CT CEV method is an appropriate experiment, but it allows a conditional model to contribute to an investment strategy not just by increasing the mean, but also by reducing the variance (which Breen, Glosten, and Jagannathan (1989) have shown to be potentially important). Moreover, it relies on an assumed risk-aversion parameter. With their $\gamma = 3$, even our unconditional prevailing mean strategy bumps against their 150% maximum investment in 12.6% of all months—mostly in the late 1990s.

Table 4 shows the resulting CEV in columns $\Delta\mathrm{U}_{\mathrm{IS}}$ and $\Delta\mathrm{U}_{\mathrm{OOS}}$. The risk-aversion parameter of 3 powerfully "amplifies" the apparent significance of predictions. Even small predictive advantage can translate into rate of return gains that are ten times as large—and either with positive and negative effects. One minor drawback is that the inference depends on the precise risk aversion parameter chosen. In addition to their assumed γ of 3, we add a column with a γ of 6.47 (7.53 for simple returns), which would have made an investor indifferent between the risk-free interest rate and the market. This changes one's inference about the economic magnitude of the models, though not in a unidirectional manner.[6]

However, even with the truncation, this utility-based mapping fails to aid four of our eight IS-significant candidate variables (**d/y**, **e/p**, **b/m**, and **all**). We shall discuss the remaining four models (**csp**, **eqis**, **ntis**, **cay3**) in further detail below.

6.1.2 Other Differences Explaining Different Perspectives

CT offer good suggestions, but we believe the reason for their better performance and therefore their more optimistic perspective lies elsewhere: they introduce two variables we did not originally consider, and they predict simple returns (in our Panel B) not log returns. Remarkably, the latter and seemingly innocuous measurement difference increases the number of IS significant models from eight to eleven, adding **tbl**, **dfy**, and **d/p**. (In our opinion, the fact that such small changes can cause differences in inference hint at the sensitivity of the models.) Yet, we shall argue that an investor should still not adopt the CT perspective.

As to new variables, CT introduce the **eqis** variable into the monthly regressions, modify the **cay** model into one that we shall dub **cay3**, and add a more recent measure, the Polk, Thompson, and Vuolteenaho (2006) **csp**. These three models rank as their best performers:

[6] CT show in equation (8) of their paper that the utility gain is roughly equal to OOS-R^2/γ. This magnification effect occurs primarily on a monthly horizons because the difference between OOS-R^2 and the ΔRMSE scales with the square root of the forecasting horizon (for small ΔRMSE, OOS-$R^2 \approx 2\Delta\mathrm{RMSE}/\mathrm{StdDev(R)}$). 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 also did repeat the CT CEV equivalent to confirm our statement.)

1. CT reasonably include **eqis**, because they do not explore annual predictions. It offers an investor 13.7 bp/month superior CEV-equivalent performance and requires not much trading (our final column). It is an equity-aggressive strategy. With $\gamma = 3$, trading based on this variable leads to the maximum permitted equity investment of 150% in 55% of all months. We have already discussed our concern with this variable as a predictor on Page 11—the fact that it relies on the good Oil Shock years, and that it has not performed well in the last thirty years. (This also comes out clearly in the monthly plot in Figure 3.) We can repeat the monthly equivalent of our summary,

eqis (CT)	All D+20	Nadir–Zenith Jul 1949 – Jul 1982: All years	Nadir–Zenith Jul 1949 – Jul 1982: w/o Oil Shock	Recent 3 Decades
ΔSSE	0.0063	0.0150	0.0073	–0.0038
ΔRMSE	0.011%	0.048%	0.026%	–0.012%

   where we use the NBER recession from Nov 1973 to Mar 1975 as the Oil Shock years. (The results would be starker if we defined it as the period over which the S&P500 declined.) This shows again that about half of **eqis**' best period performance was the Oil Shock recession, a two-year performance that is larger than **eqis**' total sample performance; and that the performance over the most recent three decades has been negative.

2. CT modify **cay** in a novel fashion. They predict the equity premium not with the linear **cay**, but with all three of its highly cointegrated ingredients. We name this novel model **cay3**. It seems to have stellar performance, both IS and OOS, gaining a market timer 20.08 bp/month superior CEV-equivalent performance, again with only modest trading. Our unreported exploration shows that the **cay** model and **cay3** models are really 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 is therefore not as well theoretically founded as the Lettau and Ludvigson (2001) **cay**, but its components are presumably known ex-ante and therefore fair game for prediction.

   However, it is an issue for trading purposes that the **cay3** model relies on data that is not immediately available. Its components are publicly released by the BEA about 1-2 months after the fact. Adding some delay to trading reduces **caya**'s performance as follows

	dRMSE_{PN}	dRMSE_{CT}	dU
Immediate Availability	–0.23 bp	+1.96 bp	+21.89 bp
One Month Delayed	–2.94 bp	–0.50 bp	+1.17 bp
Two Months Delayed	–3.58 bp	+0.15 bp	+4.27 bp

Moreover, it is again worrisome *when* the **cay3** model predicts well. Figure 3 shows that although **cay3** failed to predict the Oil Shock decline, about one third of its performance stems from the recovery period right *after* the Oil Shock.

cay3 (CT)	All	Nadir–Zenith Jan 1973 – Jan 1976		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.0060	0.0124	0.0086	–0.0025
ΔRMSE	0.017%	0.279%	0.514%	–0.008%

In any case, **cay3** has not performed well for the last three decades. Finlly, **cay3** has also excused itself from prediction over most of the most recent decade, being truncated to zero. Thus, it is of no current usefulness even to an investor who believes in it.

3. CT introduced **csp** into the set of models. Developed by Polk, Thompson, and Vuolteenaho (2006), it is the premium of high-beta over low-beta stocks. Panel B shows that it offers a market timer 6.31 bp/month superior CEV-equivalent performance. Our plot in Figure 3 shows when and how this variable actually predicts, and diagnoses potential concerns:

   (a) **csp** had good performance from September 1965 to March 1980. Unfortunately, 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. All of **csp**'s superior performance has occurred since August 2001. Although it is commendable that it has performed well late rather than early, some good performance over its first 45 years would have made us deem this variable more reliable.

   (b) The **csp** model excuses itself in 43.6% of all months from making a prediction, instead truncating the forecast to zero.

   (c) The main reason why the CT truncated version performs better than the plain OLS version is that it excuses the **csp** variable from predicting (poorly) from July 1957 through January 1963 (with interruptions). The CT truncations make little difference thereafter. It is the treatment of these specific 66 months that would make an investor either believe in superior positive or inferior outright negative performance for **csp**. We do not understand why this particular 66 month period is crucial.

   (d) The Oil Shock recession can account for $0.0132-0.0088 \approx 0.0044$ of **csp**'s performance, a number that is not only one-third of its best period performance, but also larger than **csp**'s full sample performance.

Repeating our earlier summaries, **cay3** had the following characteristics:

csp (CT)	All	Nadir–Zenith May 1965 – Jul 1982		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.0037	0.0132	0.0088	–0.0005
ΔRMSE	0.0079%	0.073%	0.059%	–0.002%

This leaves **ntis** and **tbl** as models with positive OOS and market timing performance.[7]

ntis offers a market timer with a gamma of 3 the CEV equivalent of 1.98 bp/month. (It reaches the maximum investment constraint in 57.2% of all months.) These 1.98 bp is likely to be offset by trading costs to turn over an additional 4.6% of the portfolio every month.[8] An investor with higher risk-aversion, who would not have been so eager to highly lever herself into the market, would have experienced a negative CEV, however.

tbl is an insignificant variable IS if we forecast log returns. If we forecast simple returns, it is statistically significant at the 9.7% level. If an investor is comfortable with this IS performance, we can proceed to the OOS evidence. It has a ΔRMSE advantage of only 0.16 (or 0.82) bp/month, but this translates into a 9.64 (or 8.76) bp/month market timing advantage. Again, the performance is Oil Shock dependent, and has disappeared after 1976. We can further diagnose its total cumulative superior performance of 0.0047 into

tbl (CT)	All	Nadir–Zenith Apr 1956 – Sep 1974		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.0047	0.0148	0.0071	–0.0051
ΔRMSE	0.008%	0.088%	0.047%	–0.016%

6.1.3 Comparing the Perspectives

Although our perspective (that these models are not good enough for actual investing) is different from CT's, we believe that we have considerable agreement.

1. We agree with their points that one can reasonably truncate the models for predictions; and that on monthly horizon, even if a small predictive ΔRMSE difference

[7] Among models that are IS insignificant, but OOS significant, none had positive performance from 1975 to today.

[8] 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 gauge, but a typical contract for a notional amount of $250,000 may cost 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.

can have relatively large consequences for a very risk-averse investor. These are good suggestions.

2. We agree that many variables in the academic literature no longer have IS significance (even at the 90% level), disqualifying them as forecasters.

3. We agree that OOS performance should not be used for primary analysis. Identification and variable selection is better left to IS regressions. We would not want to invest based on a variable which is IS insignificant, regardless of OOS performance. Instead, OOS performance is an important regression *diagnostic*, but only conditional on the model being IS significant. Therefore, we probably agree that because OOS power is only relevant in IS statistically significant regressions, the statistical power of the OOS tests is often quite good.

4. We probably agree that 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.)

5. We probably agree that an investor would have had to have known *ex-ante* which of the models would have held up.

6. We probably agree that even after the CT suggestions, many models earned negative certainty equivalents.

7. We probably agree that none of the models had superior performance over the last three decades, although we would be relatively more inclined to attribute this to unstable and therefore now useless models.

In sum, we believe an investor should be aware of the issues pointed out in the CT paper and our own paper—and then judge whether these prediction models are sufficiently reliable for investment purposes.

7 Alternative Specifications

We also explored some alternative models and specification which have been proposed as improvements over the simple regression specifications.

7.1 Longer-Memory Dividend and Earnings Ratios

Table 5 considers dividend-price ratios, and earnings-price ratios, and dividend-earnings ratios with memory (which simply means that we add multiple year dividends or earnings). The table is an excerpt from a complete set of 1-year, 5-year, and 10-year dividend-price ratios, earnings-price ratios, and dividend-earnings ratios. (That is, we tried all 90 possible model combinations.) The table contains *all* IS significant specifications from our monthly regressions begin forecasting in 1965, our monthly, annual, and five-yearly forecasts begin either in 1902 or 1965.

Even though there were more combinations of dividend-earnings ratios than either dividend-price or earnings-price ratio, *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 performance, all of which were statistically significant. (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 3 decades. We leave it to the reader to decide whether they believe that real-world investors would have been able to choose the right models for prediction.

7.2 Alternative Estimation 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. (The exceptions are **ntis**, **ltr**, and **dfy**.) Our previously reported statistics take this into account, because we bootstrapped for significance levels mimicking the IS autocorrelation of each independent variable. However, the Stambaugh (1999) coefficient correction is a more powerful test in non-asymptotic samples. Lewellen (2004) and Campbell and Yogo (2003) further improve on the estimation technique by assuming different boundary behavior. (An alternative technique to account for autocorrelation in dividend ratios in Goyal and Welch (2003) is structural, not statistical, and thus can apply only to the dividend price ratio.) This section therefore explores equity premium forecasts using these corrected coefficients.

In the upper panel in Table 6, 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 correction is surprisingly powerful IS. Given our bootstrapped critical rejection levels under the NULL hypothesis, it was able to identify eight (rather than just three) ALTERNATIVE models as different from zero. In six of them, it even imputed significance in each and every of our 10,000 bootstraps! The Lewellen coefficient is often dramatically different from the OLS coefficients, resulting in negative $\overline{R}^2$ even among its IS significant variable estimations.

Unfortunately, neither the Stambaugh nor the Lewellen technique manage 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—although it is our only variable that had non-negative OOS performance over the last three decades.

e/p	All	Nadir–Zenith Sep 1974–Feb 2003		Recent 3
	D+20	All years	w/o Oil Shock	Decades
ΔSSE	0.0015	0.0032	0.0024	0.0013
ΔRMSE	0.004%	0.010%	0.008%	0.004%

8 Encompassing Tests

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 $\Delta\text{RMSE}^{\star}$. However, we can also presume that investors would not have known the optimal *ex-post* λ. This means that they would have computed λ based on 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 $\Delta\text{RMSE}^{\star r}$.[9]

Table 7 shows the results of encompassing forecast estimates. Panel A predicts annual equity premia. Necessarily, all ex-post λ combinations have positive $\Delta\text{RMSE}^{\star}$—but almost all rolling λ combinations have negative $\Delta\text{RMSE}^{\star r}$. The exceptions are **d/e**, and **dfy**, **all**, and **caya** in some but not all specifications. **d/e** and **dfy** can immediately be excluded, because their optimal λ is negative. This leaves **all** and **caya**. Again, not reported, these could not outperform over the most recent three decades.

In the monthly rolling encompassing tests, only **svar** and **d/e** (in one specification) are positive, neither with a positive λ.

In sum, "learned shrinking" would not have improved any of our models to the point where we would have expected them to outperform—quite the opposite.

[9] For the first three observations, we presume perfect optimal foresight, resulting in the minimum ΔRMSE. This tilts the rolling statistic in favor of no inferior performance. The results remain the same if we use reasonable alternatives for when we begin using historical rolling λ estimates and what λ we use until this point.

9 Other Literature

Our paper is of course not the first to explore equity premium prediction or OOS tests.[10] Many of the earlier OOS tests have focused on the dividend ratios. The models that have most influenced our paper were the following:

- Fama and French (1988) interpreted the OOS performance of dividend ratios to have been a success. Our paper comes to the opposite conclusion primarily because the sample period has been extended.
- Bossaerts and Hillion (1999) interpreted the OOS performance of the dividend yield (not dividend price ratio) to be a failure, too. However, they relied 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 are difficult to extrapolate.
- 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 some detail on annual horizons—and in more detail than our own paper. Our current paper expands the set of variables and data periodicities to be comprehensive, explores different techniques, and considers issues of power.

OOS tests have also been used in the context of other models. For example, Lettau and Ludvigson (2001) run rolling OOS regressions—but not in the same spirit as our paper: the construction of their CAY variable itself relies on ex-post data. 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 paper pursues.

There are at least three studies in which authors seek to explore a more comprehensive set of variables:

- Pesaran and Timmerman (1995) (and others) have pointed 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. They explore 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, beginning in 1954 and ends (by necessity) twelve years ago in 1992. They conclude

[10] There are many other papers that have critiqued predictive regressions on other grounds than OOS. In particular, the use of *dividend ratios* has been critiqued in many other papers (see, e.g., Goetzmann and Jorion (1993) and Ang and Bekaert (2003)). A number of papers have also documented low in-sample power (e.g., see Goetzmann and Jorion (1993), Nelson and Kim (1993), Torous and Valkanov (2000), and Valkanov (2003)). Apologies to everyone whose paper we omit to cite here—the literature is voluminous.

that investors could have succeeded, especially in the volatile periods of the 1970s. 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 paper shows that the Oil Shock experience generally is almost unique in making many predictive variables seem to outperform. Still, even including the two-year Oil Shock period in the sample, the overall OOS performance of our ALTERNATIVE models is typically poor.

- Ferson, Sarkissian, and Simin (2003) is not exactly based on OOS regressions, but it is interested in a closely related issue—spurious regressions and data mining in the presence of serially correlated independent variables. They suggest increasing the critical t value of the in-sample regression. The paper 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 (2003). 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 out-of-sample forecasting power.

The above papers disagree with the general tenet of predictability, but they appear to be in the minority. Still, bits and pieces of evidence we report have surfaced elsewhere, and some authors working with the data may already know what variable works and where it does not work—but this is not easy to systematically determine for a reader of this literature. The general literature tenet remains that the empirical evidence and professional consensus is generally supportive of predictability. This is why we believe that it was 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. This alone can explain some otherwise startling discrepancies in the literature.

Rapach and Wohar (2004) is perhaps closest to our paper. It is also fairly recent, also fairly comprehensive, and also explores out-of-sample performance for a number of variables. We come to many similar conclusions. Their study ends in 1999, while our data end in 2004—a fairly dramatic five years. Moreover, our study focuses more on diagnosis of weaknesses, than just on detection.[11,12] Moreover, our paper also focuses on pointing out the performance *after* the Oil Shock.

[11] A particularly interesting factoid applies to **eqis** (the share of equity issued by corporations, see below), which they (like us) deem to be the best predictor for annual equity premium forecasting. If we add 2000–2004, the performance of **eqis** deteriorates to make it insignificant based on MSE-T statistic.

[12] 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 checked 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.

10 Conclusion

Findings Our paper systematically investigates the IS and OOS performance of linear regressions to predict the equity premium with variables popular in academic. We believe that the evidence suggests none of the academic models we reexamined warrants a strong investment endorsement. Most models not only cannot beat the unconditional benchmark, but also outright underperform it. Our conclusions can be regarded as being conservative because we do not conduct a true OOS test—we include the same data that were used to establish the models. We also ignore the question of how an investor would have known which of the many models we considered would ultimately have worked.

Our plots help diagnose when and why models failed. They show why, standing in 2005 (or earlier), popular academic models have not worked IS or OOS or both. If we exclude the two consecutive years of the 1973 Oil Shock, most models would have performed even worse. Most importantly, no model has been reasonably robust, performing similarly over different subperiods—and specifically, no model has done well over the most recent three decades. For the Consequently, for the very few models that were significant in the full sample (such as **eqis**), an investor would have to have faith not just in the results from the most thirty years, but in the full-period long-run—and have faith that the underlying model has not changed.

Our findings are not cherry-picking. On rare occasion, a choice of sample period, data frequency, and method can lead to IS statistical significance for a variable, more rarely both IS and OOS statistical significance, but we believe such rare performance is not robust evidence of predictability.[13] Our findings are also not driven by power that is weaker than those of earlier papers. Our findings are not driven by a few outlier years. Our findings do not disappear if we use different earnings and dividends definition and corrections for the time-series properties of the independent variable. The models perform no better in an encompassing OOS tests, provided we use a test that does not rely on *ex-post* information.

In sum, by assuming that the equity premium is "like it always has been," an investor would have predicted just as well. (To draw this conclusion, we relied not only on the printed tables in our paper, but on a much larger set of tables that explored combinations of alternative data definitions, data frequencies, time periods, econometric specifications, etc).[14]

[13] Our earlier draft was critiqued by referees for coming to opposite conclusions as those from other circulating or forthcoming papers, specifically Guo (2006) and Rapach and Wohar (2004). We replicated these papers, and found that differences were due to sample period or year-weighting.

[14] The tables in this paper have been distilled from a larger set of 52 table panels on around sixty pages, which are available from our website—and on which we sometimes draw in our text description of results. These tables also explore a set of monthly price changes from 1871 to 1926 from Robert Shiller's website to compute monthly percent price changes after 1871, which were the main monthly tables in our earlier draft.

Directions We can speculate why the models generally performed so poorly. It is probably instability in the underlying models. This instability could be in the linear prediction model or in the process driving the independent variable.[15] The evidence in the most recent thirty years speaks especially strongly against all models explored.

If we now explored more variables and/or more sophisticated models (e.g., through structural shifts or Kalman filters), then we face the issue of specification search even more strongly. Some of these models are bound to work both IS or OOS by pure chance. Thus, researchers should then need to wait for more and new OOS data to become available in order to test such new variables or more sophisticated models. Maybe theory is our best hope—an argument along the line of Cochrane (1997)'s observation that the dividend yield must predict future returns eventually if it fails to predict dividend growth.

Our paper is simple, but we believe its implications are not. The belief that the state variables which we explored in our paper 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 of 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 paper suggests only that our profession has yet to find a variable that has meaningful robust empirical equity premium forecasting power, both IS and OOS, at least from the perspective of a real-world investor. We hope that the simplicity of our approach strengthens the credibility of our evidence.

We close by paraphrasing Mark Twain's famous line, admittedly with some tongue-in-cheek:

The rumors of the predictability of the
equity premium are greatly exaggerated.

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_2004.html`.

Bureau of Labor Statistics Webpage: `http://www.bls.gov/cpi/`

[15] Not reported, we can statistically reject regression model stability for all variables we examined in our monthly tests, and for almost all variable in our annual tests [except **dfs**, **infl**, and **cayp**] by using the CUSUMQ test, which is known to be fairly weak.

Martin Lettau's Webpage: (**cay**), http://pages.stern.nyu.edu/∼mlettau/.

Jeff Wurgler's Webpage: (**eqis**), http://pages.stern.nyu.edu/∼jwurgler/

Figure 1: Annual Performance of In-Sample Insignificant Predictors

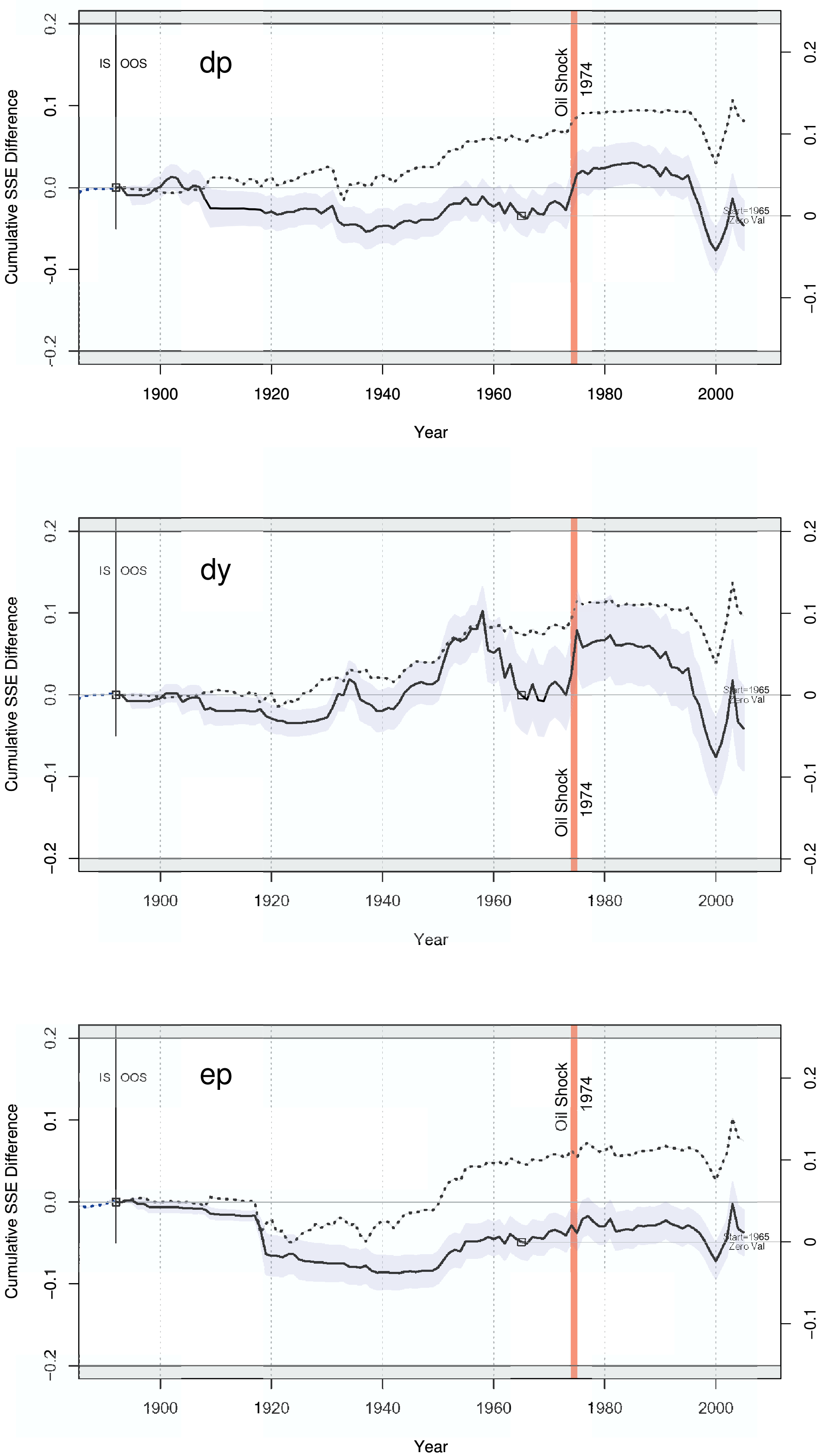


Figure 1: con't

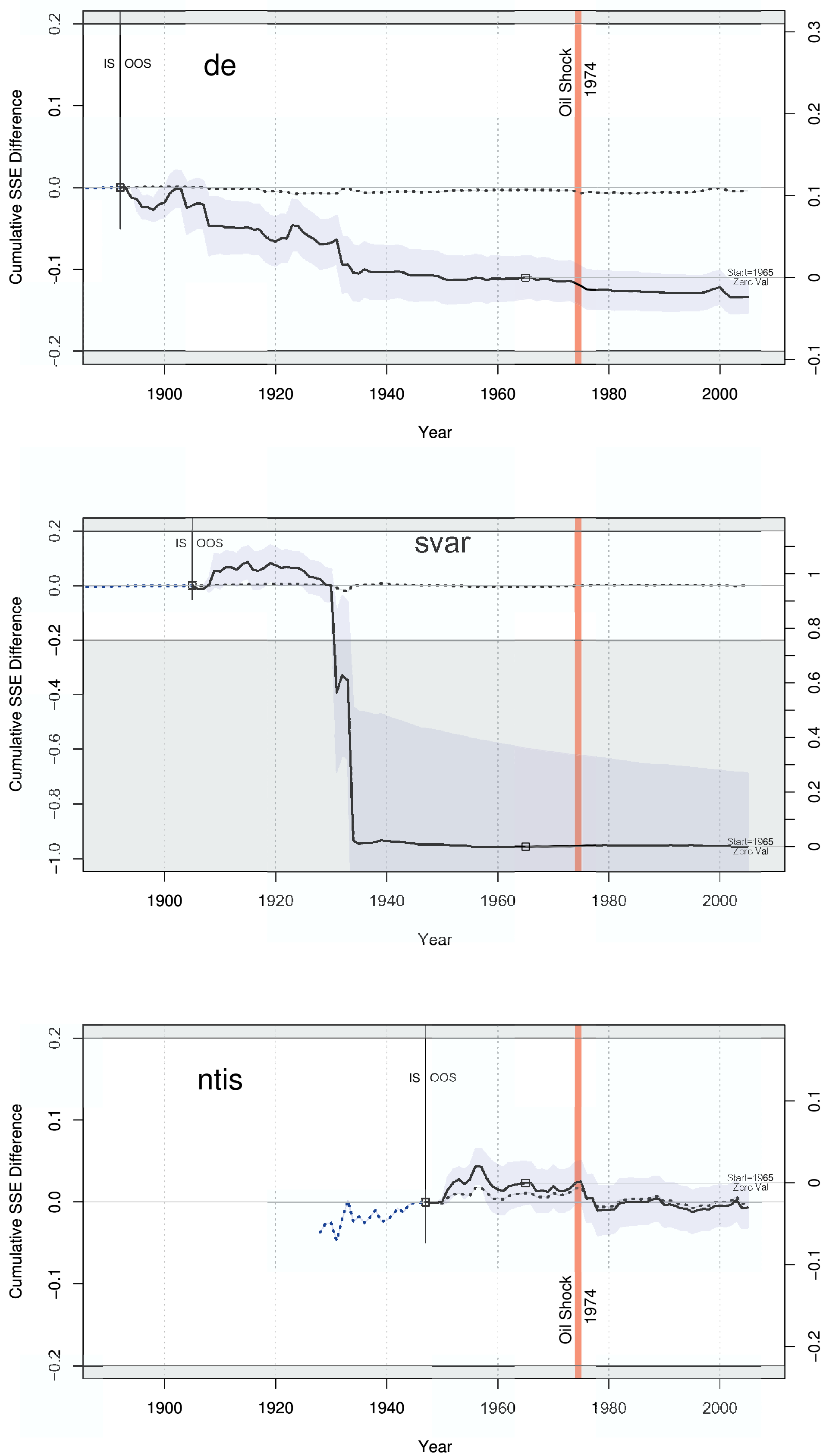


Figure 1: con't

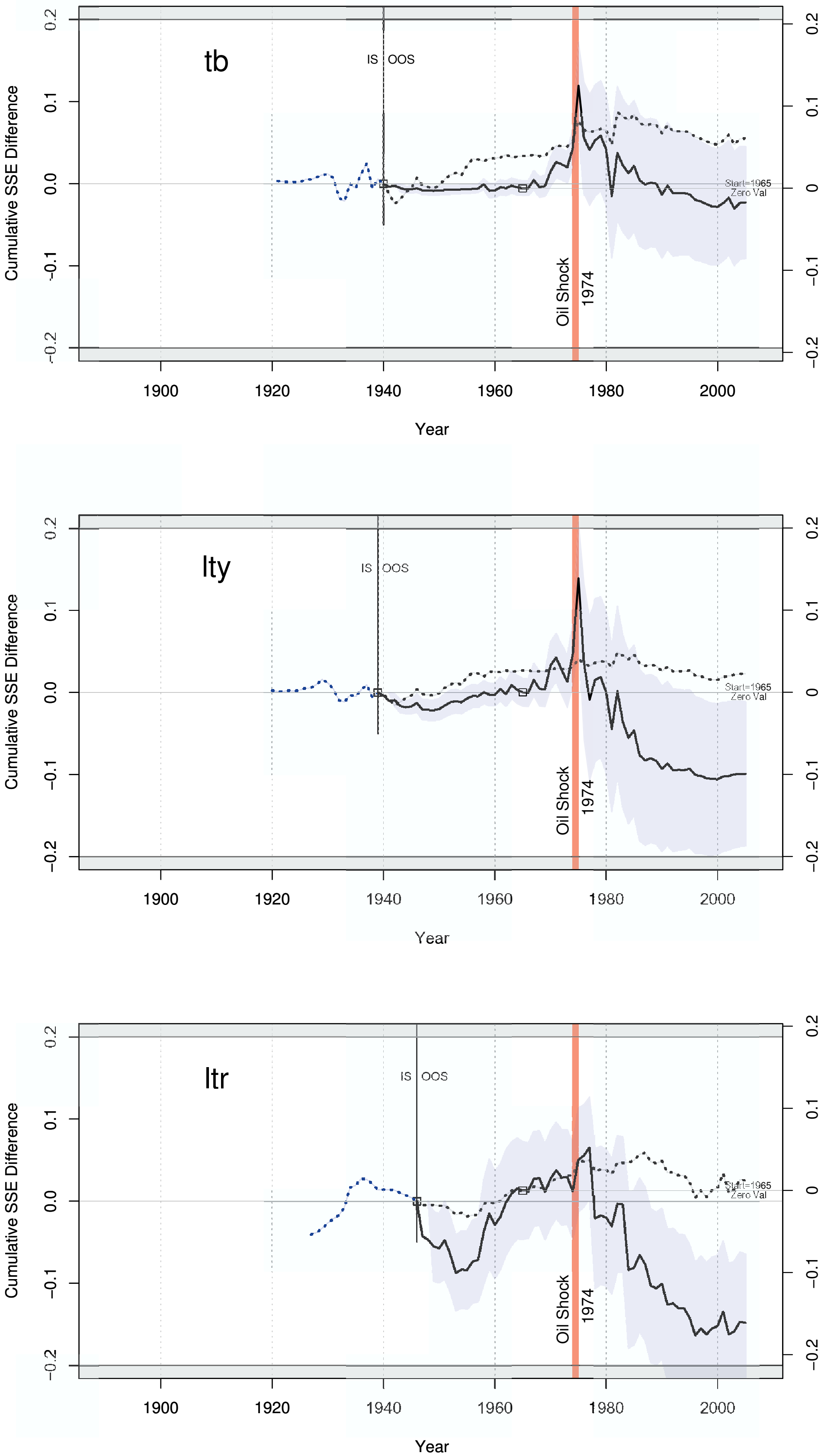
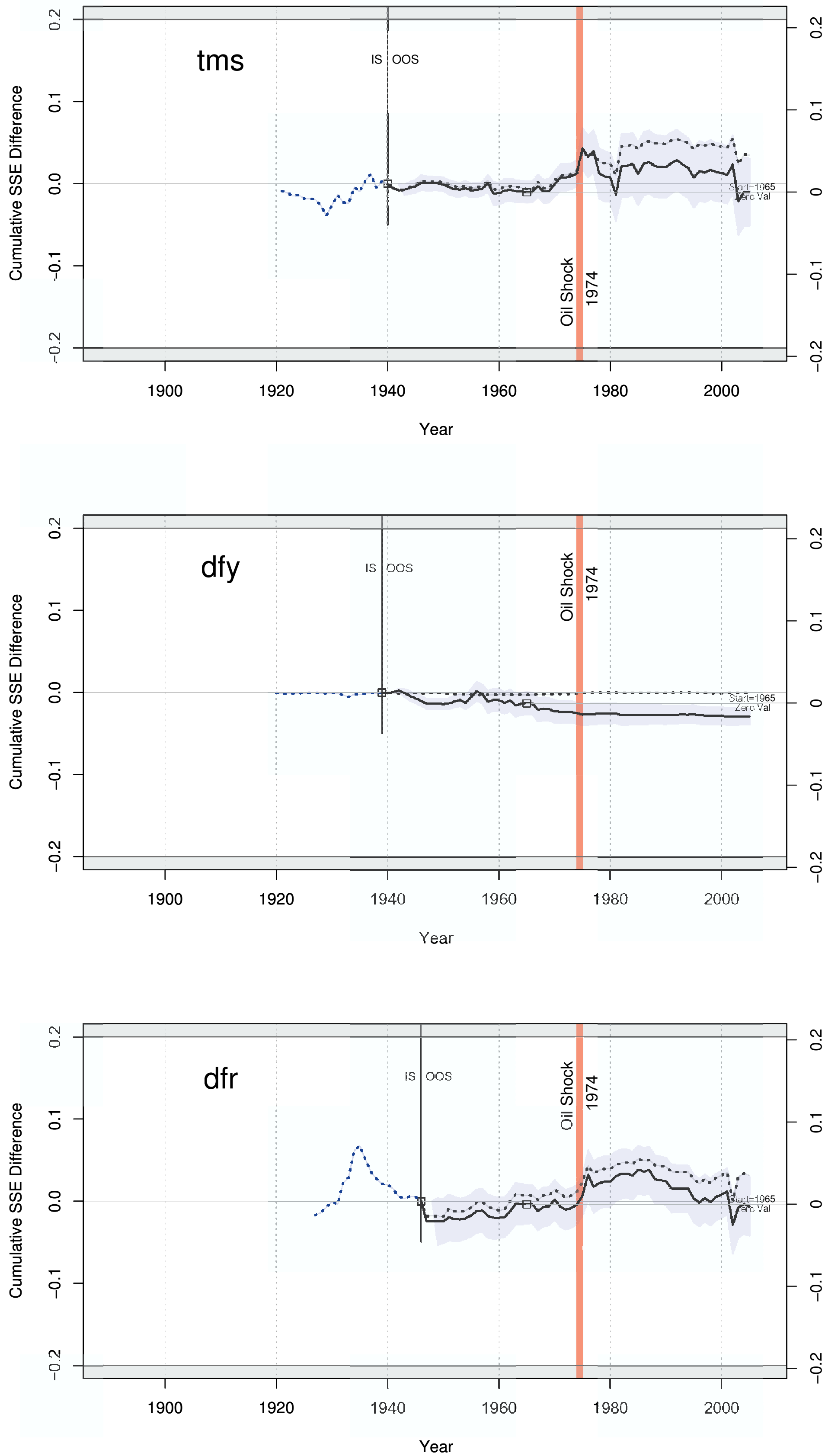
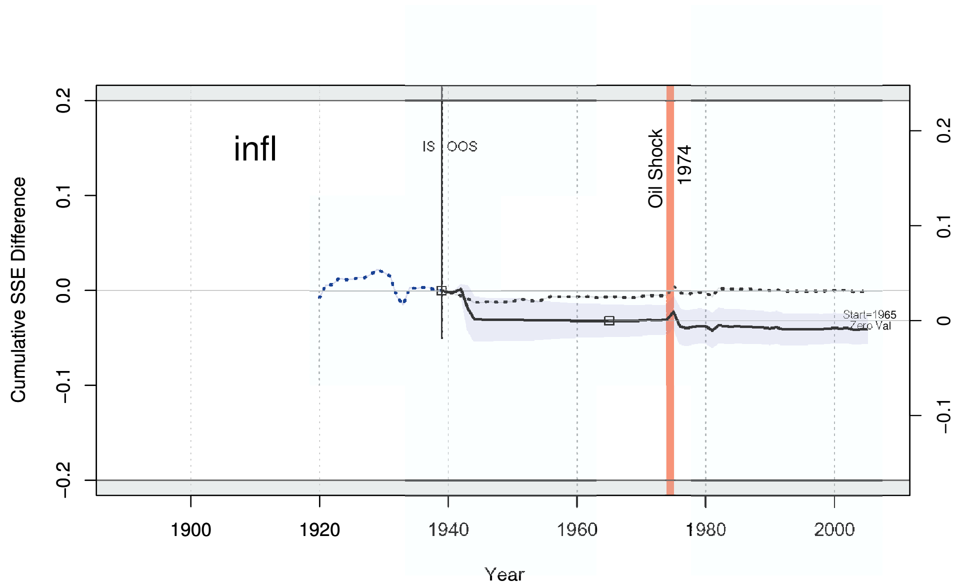


Figure 1: con't





Explanation: These figures plot the IS and OOS performance of annual predictive regressions. These are the cumulative squared prediction errors of the prevailing mean NULL minus the cumulative squared prediction error of the predictive variable from a linear historical regression. For the OOS graph, the NULL is the prevailing mean. The IS prediction is dotted, the OOS 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 2: Annual Performance of In-Sample Significant Predictors

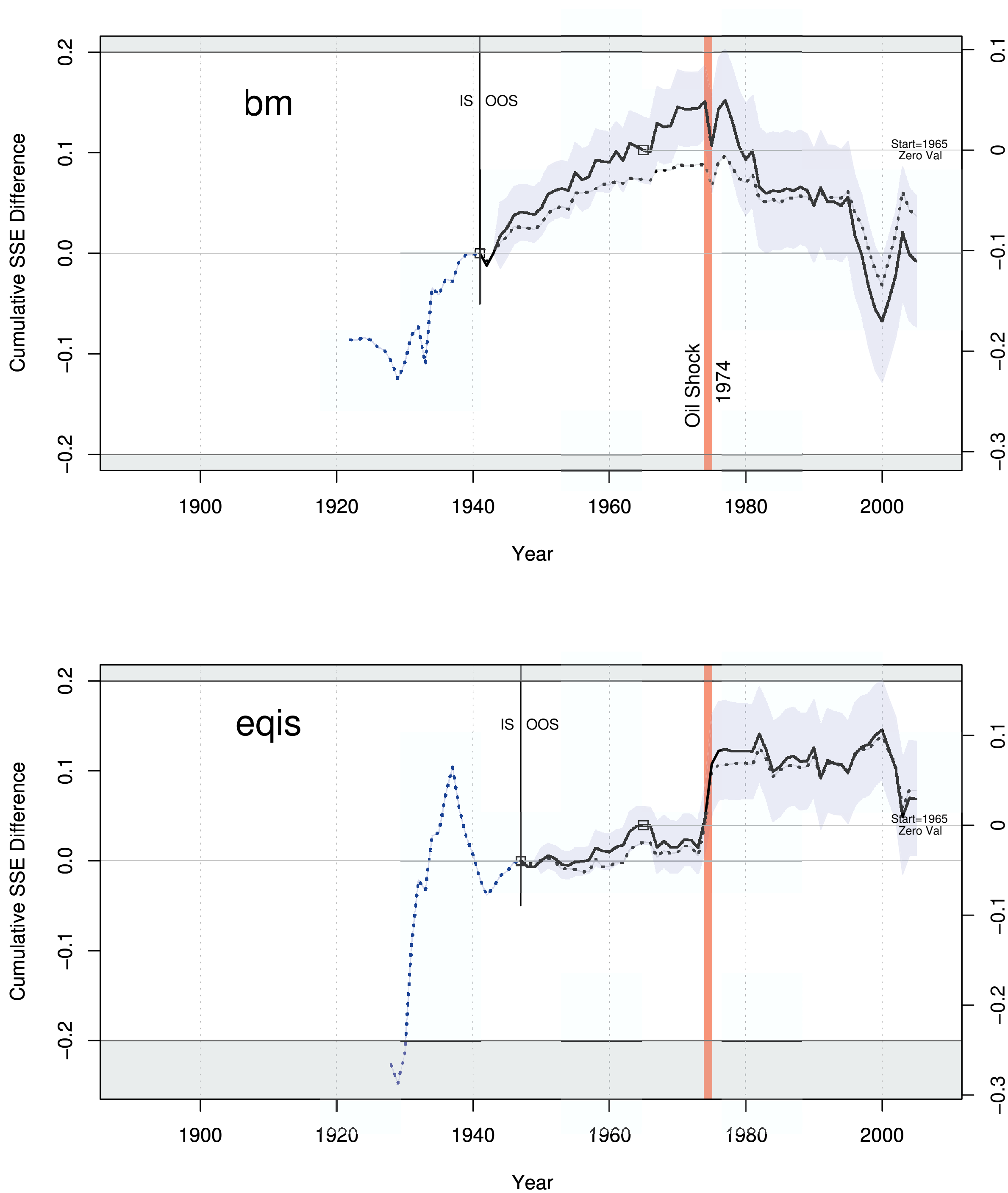


Figure 2: con't

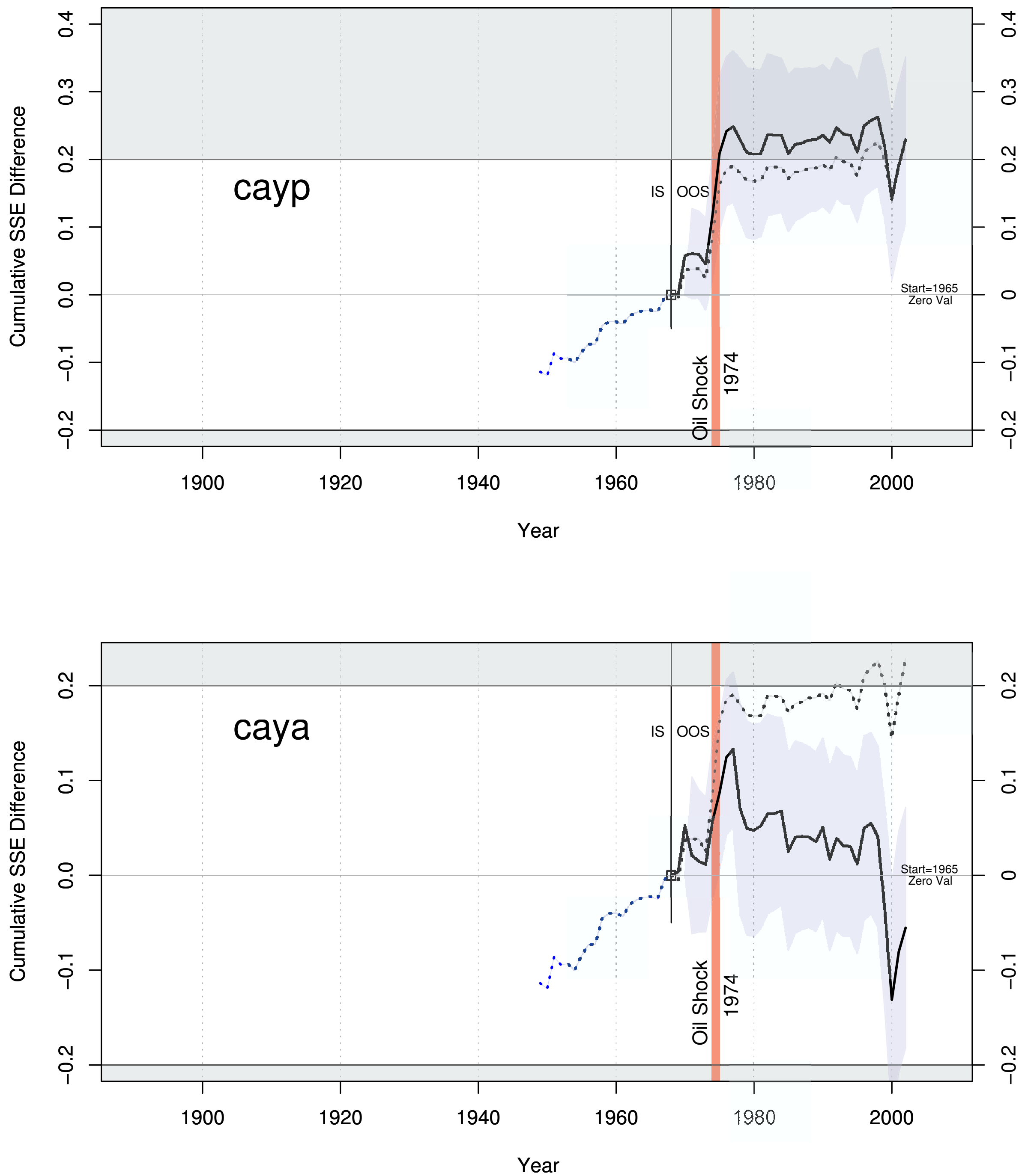
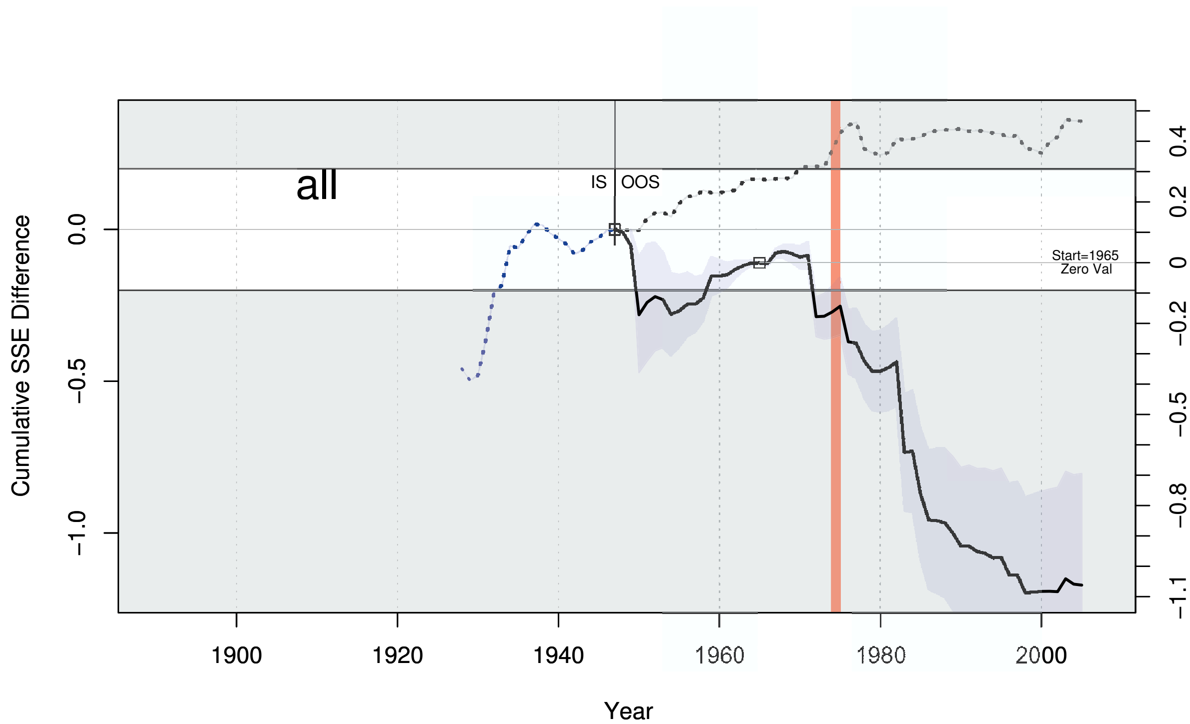
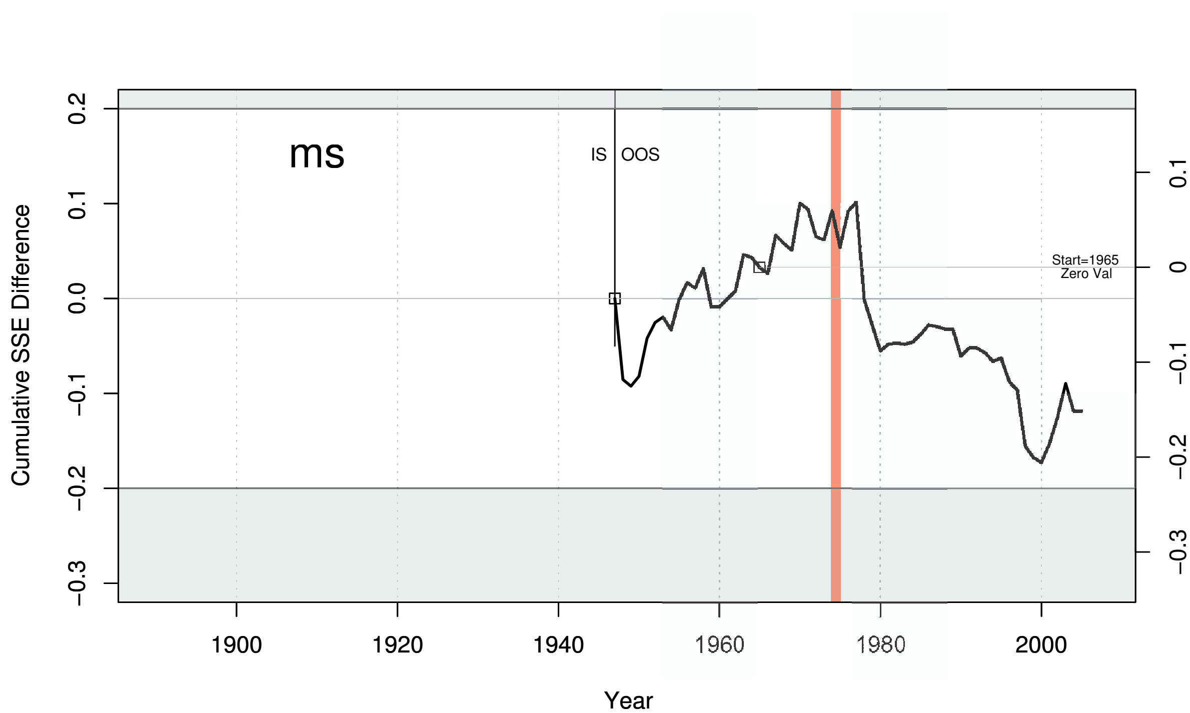


Figure 2: con't





Explanation: See Figure 1.

Figure 3: Monthly Performance of In-Sample Significant Predictors

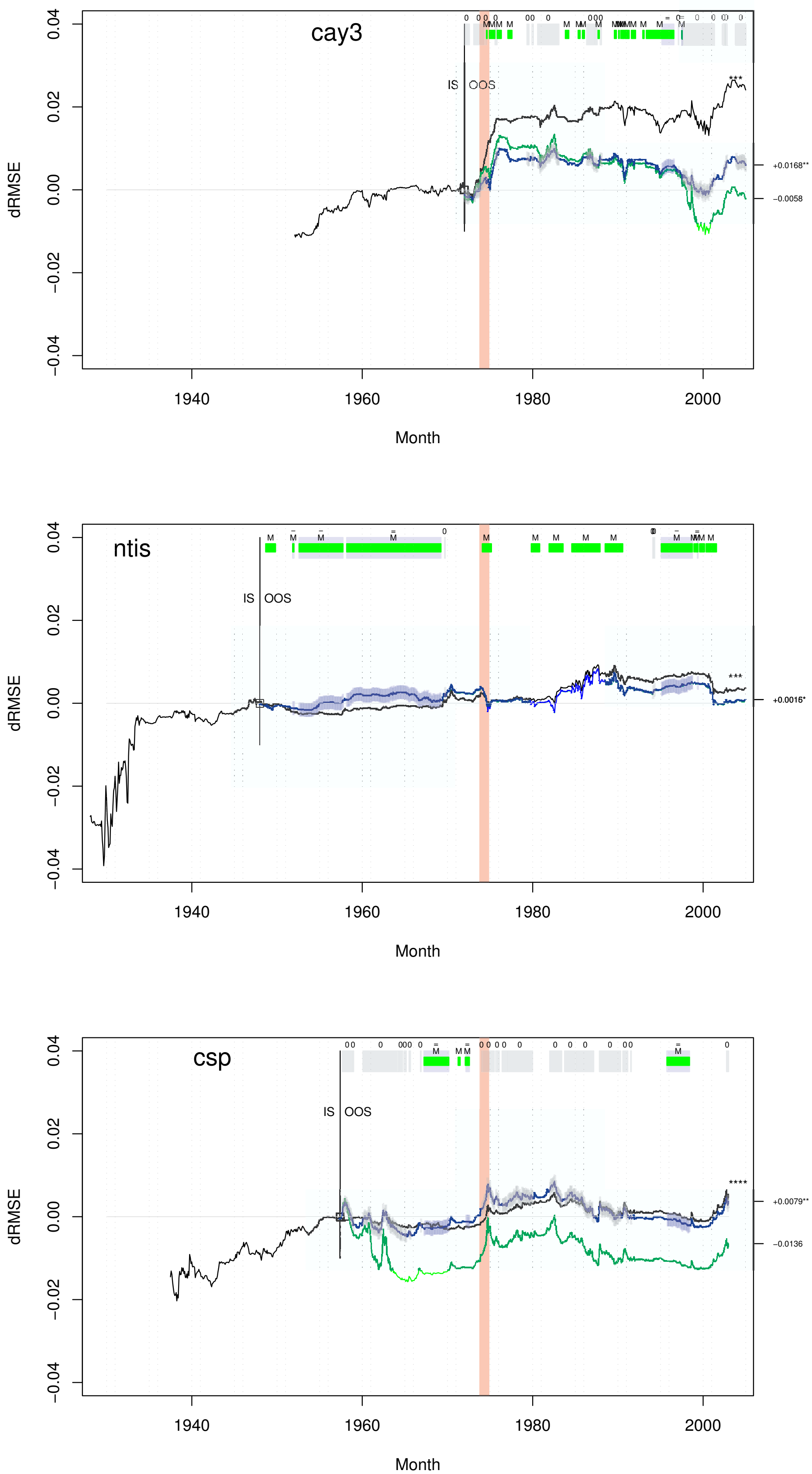
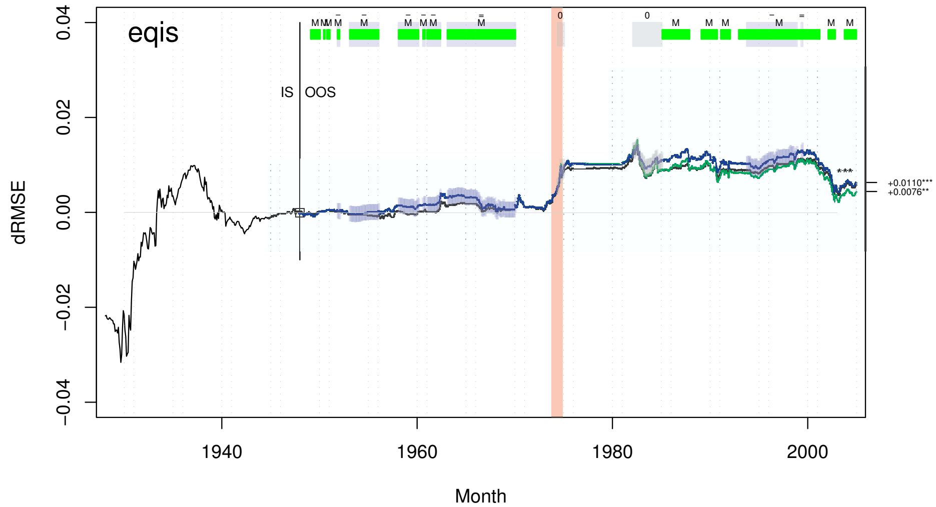
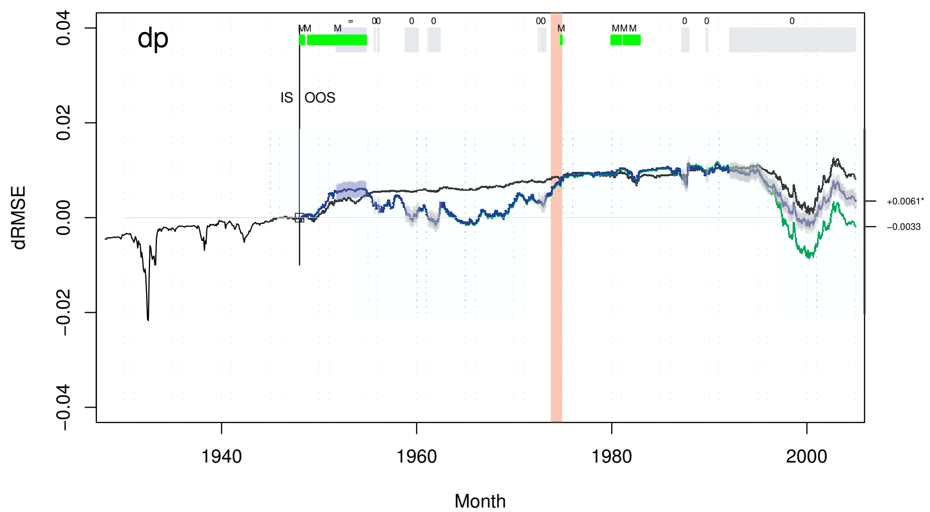
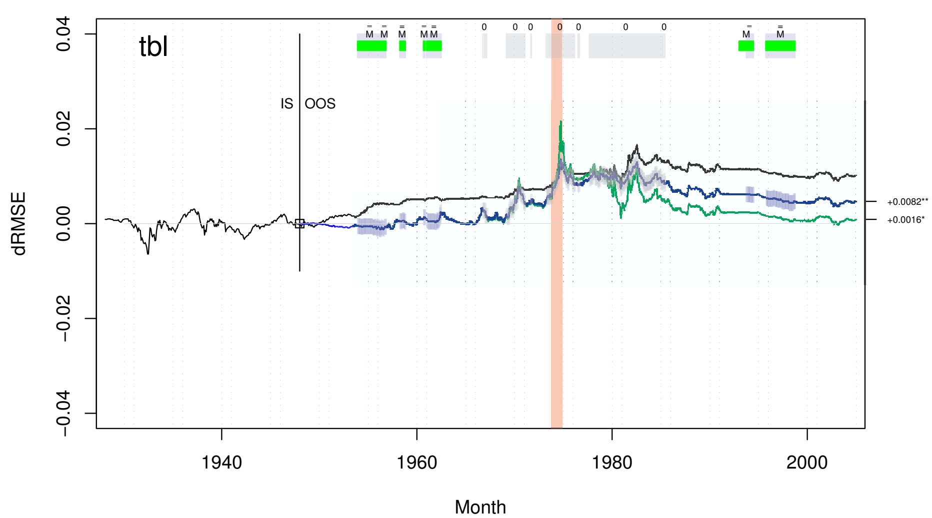


Figure 3: cont'd







Explanation: These figures are the analogs of figure 1, plotting the IS and OOS performance of the named model. However, they use monthly data. The Campbell-Thompson (2005) (CT) performance is plotted in blue, the plain model performance is plotted in green. The bars at the top indicate when the CT model makes a non-linear prediction—a "0" indicates truncation at 0, an "=" indicates a wrong sign coefficient, in which case CT revert to the unconditional model. (When the CT risk-averse investor would purchase equities worth 150% of his wealth, the maximum permitted, it is marked by a "=" in the figure.) In addition to the bars at the top, we have also marked this by fattening the CT OOS prediction line. In all cases plotted here, the CT performance is between the IS and the plain OOS performance. The Oil Shock is marked by a red vertical line.

Table 1: Forecasts at Annual Frequency

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at the annual frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. Stock return is price changes, including dividends, of S&P500. Panel A presents the results for insignificant predictors while Panel B presents the results for significant predictors. The data period for each variable is indicated next to it. The column heading 'D+20' in Panel A begins forecast 20 years after the sample date while the column heading '1965-' in Panel A begins forecast in 1965. All numbers, except $\overline{R}^2$ and power, are in percent per year. A star next to $\overline{R}^2$ denotes significance of the in-sample regression (as measured by empirical F-statistic). 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 column 'IS for OOS' in Panel B gives the ΔRMSE of IS errors for OOS period. MSE-F is 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 **ms** model are not calculated). 'Pwr' is the power of ΔRMSE and is calculated as the fraction of draws where the simulated ΔRMSE is greater than the empirically calculated 95% critical value and is reported in percent. The two numbers under the power column are power for all simulations and simulations that are found to be in-sample significant at the 95% level. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

Panel A: Insignificant in-sample predictors

	Variable	Full Data				1927–2004 Data	
		Data	IS	OOS		IS	OOS
			$\overline{R}^2$	ΔRMSE D+20	ΔRMSE 1965–	$\overline{R}^2$	ΔRMSE 1965–
d/p	Dividend Price Ratio	1872–2004	0.47	−0.1092	−0.0908	1.60	0.0558
d/y	Dividend Yield	1872–2004	0.89	−0.0971	−0.3162	See Panel B	
e/p	Earning Price Ratio	1872–2004	1.00	−0.0886	0.0899	See Panel B	
d/e	Dividend Payout Ratio	1872–2004	−0.75	−0.3140	−0.1846	−1.24	−0.5659
svar	Stock Variance	1885–2004	−0.76	−2.3405	0.0104	−1.33	−0.0715
ntis	Net Equity Expansion	1927–2004	−0.03	−0.0352	−0.2303	−0.03	−0.2303
tbl	T-Bill Rate	1920–2004	0.57	−0.1083	−0.1318	0.37	−0.5473
lty	Long Term Yield	1919–2004	−0.53	−0.4638	−0.7499	−0.86	−1.0074
ltr	Long Term Return	1926–2004	1.00	−0.7696	−1.2016	0.93	−1.0396
tms	Term Spread	1920–2004	0.30	−0.0488	−0.0008	1.06	0.0315
dfy	Default Yield Spread	1919–2004	−1.20	−0.1376	−0.1249	−1.33	−0.0986
dfr	Default Return Spread	1926–2004	0.38	−0.0330	−0.0194	0.30	−0.0100
infl	Inflation	1919–2004	−0.98	−0.1939	−0.0714	−1.05	−0.4168

Panel B: Significant in-sample predictors

			IS		Forecasts begin 20 years after sample data				Forecasts begin 1965			
					IS for OOS	OOS			IS for OOS	OOS		
	Variable	Data	$\overline{R}^2$	ΔRMSE	ΔRMSE	ΔRMSE	MSE-F	Pwr	ΔRMSE	ΔRMSE	MSE-F	Pwr
b/m	Book to Market	1921–2004	3.01*	0.3987*	0.1854	–0.0383	–0.31	41 (**68**,9)	–0.2789	–0.8275	–3.78	38 (**60**,12)
eqis	Pct Equity Issuing	1927–2004	9.62***	1.0641***	0.4332*	0.3765	2.83**	75 (**85**,15)	0.4591*	0.2313	1.18*	68 (**78**,15)
cayp	Cnsmptn, Wlth, Incme	1948–2001	24.89***	2.2188***	2.3357***	2.2418	11.83***	88 (**90**,12)	2.3357***	2.2418	11.83***	88 (**90**,12)
all	Kitchen Sink	1927–2004	15.61**	3.0042	2.0892	–5.4030	–25.59	— (—,—)	1.5728	–6.8113	–20.26	— (—,—)
caya	Cnsmptn, Wlth, Incme	1948–2001	—	—	—	–0.4958	–1.99	— (—,—)	—	–0.4958	–1.99	— (—,—)
ms	Model Selection	1927–2004	—	—	—	–0.6272	–4.31	— (—,—)	—	–1.1309	–5.06	— (—,—)

			IS		Forecasts begin 1965			
					IS for OOS	OOS		
	Variable	Data	$\overline{R}^2$	ΔRMSE	ΔRMSE	ΔRMSE	MSE-F	Pwr
d/y	Dividend Yield	1927–2004	2.65*	0.3799*	0.2421	–0.3121	–1.51	30 (**72**,6)
e/p	Earning Price Ratio	1927–2004	3.01*	0.4147*	0.1635	–0.0861	–0.42	38 (**64**,11)
b/m	Book to Market	1927–2004	3.97*	0.5079*	–0.4097	–1.3070	–5.79	42 (**61**,13)
eqis	Pct Equity Issuing	1927–2004	9.62***	1.0641***	0.4591*	0.2313	1.18*	68 (**78**,15)
cayp	Cnsmptn, Wlth, Incme	1948–2001	24.89***	2.2188***	2.3357***	2.2418	11.83***	88 (**90**,12)
all	Kitchen Sink	1927–2004	15.61**	3.0042	1.5728	–6.8113	–20.26	— (—,—)
caya	Cnsmptn, Wlth, Incme	1948–2001	—	—	—	–0.4958	–1.99	— (—,—)
ms	Model Selection	1927–2004	—	—	—	–1.1309	–5.06	— (—,—)

Table 2: Forecasts at 5-year Frequency

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at the 5-year frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. Stock return is price changes, including dividends, of S&P500. Panel A presents the results for insignificant predictors while Panel B presents the results for significant predictors. The data period for each variable is indicated next to it. The column heading 'D+20' in Panel A begins forecast 20 years after the sample date while the column heading '1965-' in Panel A begins forecast in 1965. All numbers, except $\overline{R}^2$ and power, are in percent per 5-years. A star next to $\overline{R}^2$ denotes significance of the in-sample regression (as measured by empirical F-statistic). 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 column 'IS for OOS' in Panel B gives the ΔRMSE of IS errors for OOS period. MSE-F is 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 **ms** model are not calculated). 'Pwr' is the power of ΔRMSE and is calculated as the fraction of draws where the simulated ΔRMSE is greater than the empirically calculated 95% critical value and is reported in percent. The two numbers under the power column are power for all simulations and simulations that are found to be in-sample significant at the 95% level. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

Panel A: Insignificant in-sample predictors

	Variable	Full Data: Data	Full Data: IS $\overline{R}^2$	Full Data: OOS ΔRMSE D+20	Full Data: OOS ΔRMSE 1965–	1927–2004 Data: IS $\overline{R}^2$	1927–2004 Data: OOS ΔRMSE 1965–
d/y	Dividend Yield	1872–2004	5.54	−0.8110	−2.6202	See Panel B	
e/p	Earning Price Ratio	1872–2004	5.55	−0.1781	−0.5398	13.41	−2.9428
d/e	Dividend Payout Ratio	1872–2004	0.50	−0.7828	0.2876	1.20	0.1171
svar	Stock Variance	1885–2004	0.50	−13.3482	0.1274	−0.71	−0.1754
b/m	Book to Market	1921–2004	9.38	−2.5234	−7.9822	12.36	−9.7919
ntis	Net Equity Expansion	1927–2004	3.15	−8.0893	0.7695*	3.15	0.7695*
tbl	T-Bill Rate	1920–2004	4.13	−2.6722	−4.5774	5.30	−9.5884
lty	Long Term Yield	1919–2004	0.10	−17.5501	−11.1243	−0.03	−17.8513
ltr	Long Term Return	1926–2004	−1.38	−1.0897	−2.7260	−1.39	−1.8780
tms	Term Spread	1920–2004	7.33	−4.4184	2.2195**	See Panel B	
dfy	Default Yield Spread	1919–2004	3.32	−9.3515	1.3001	0.77	0.4996
dfr	Default Return Spread	1926–2004	−1.15	−0.6178	0.0299	−1.09	0.0309
infl	Inflation	1919–2004	−1.19	−1.7413	−0.8830	−1.29	−1.9284

Panel B: Significant in-sample predictors

	Variable	Data	IS		Forecasts begin 20 years after sample data				Forecasts begin 1965			
			IS		IS for OOS	OOS			IS for OOS	OOS		
			$\overline{R}^2$	ΔRMSE	ΔRMSE	ΔRMSE	MSE-F	Pwr	ΔRMSE	ΔRMSE	MSE-F	Pwr
d/p	Dividend Price Ratio	1872–2004	9.78*	1.9891*	2.7133*	–0.1672	–0.92*	22 (**67**,6)	1.1935	–3.9423	–7.34	20 (**51**,9)
eqis	Pct Equity Issuing	1927–2004	10.29*	2.3305*	–0.2250	0.1219	0.34	30 (**78**,8)	0.6202	–0.2082	–0.41	28 (**70**,9)
cayp	Cnsmptn, Wlth, Incme	1948–2001	36.04***	6.8601***	9.0035***	8.4030	17.22***	61 (**81**,15)	9.0035***	8.4030	17.22***	61 (**81**,15)
all	Kitchen Sink	1927–2004	42.62***	12.0932	10.4590	–36.4433	–39.69	— (—,—)	9.6415	–31.2312	–25.61	— (—,—)
caya	Cnsmptn, Wlth, Incme	1948–2001	—	—	—	2.0592	3.22**	— (—,—)	—	2.0592	3.22**	— (—,—)
ms	Model Selection	1927–2004	—	—	—	–17.3081	–28.93	— (—,—)	—	–9.2420	–13.01	— (—,—)

	Variable	Data	IS		Forecasts begin 1965			
			IS		IS for OOS	OOS		
			$\overline{R}^2$	ΔRMSE	ΔRMSE	ΔRMSE	MSE-F	Pwr
d/p	Dividend Price Ratio	1927–2004	20.36*	4.4606*	2.0097	–2.1836	–3.97	29 (**61**,12)
d/y	Dividend Yield	1927–2004	14.20*	3.1428	2.3725	–1.6305	–3.03	24 (**74**,8)
eqis	Pct Equity Issuing	1927–2004	10.29*	2.3305*	0.6202	–0.2082	–0.41	28 (**70**,9)
tms	Term Spread	1927–2004	11.89*	2.6607*	4.2629**	2.5598	5.68**	11 (**64**,4)
cayp	Cnsmptn, Wlth, Incme	1948–2001	36.04***	6.8601***	9.0035***	8.4030	17.22***	61 (**81**,15)
all	Kitchen Sink	1927–2004	42.62***	12.0932	9.6415	–31.2312	–25.61	— (—,—)
caya	Cnsmptn, Wlth, Incme	1948–2001	—	—	—	2.0592	3.22**	— (—,—)
ms	Model Selection	1927–2004	—	—	—	–9.2420	–13.01	— (—,—)

Table 3: Forecasts at Monthly Frequency

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at the monthly frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. Stock return is price changes, including dividends, of S&P500. Panel A presents the results for insignificant predictors while Panel B presents the results for significant predictors. The data period for each variable is indicated next to it. The column heading 'D+20' in Panel A begins forecast 20 years after the sample date while the column heading '1965-' in Panel A begins forecast in 1965. All numbers, except $\overline{R}^2$ and power, are in percent per month. A star next to $\overline{R}^2$ denotes significance of the in-sample regression (as measured by empirical F-statistic). 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 column 'IS for OOS' in Panel B gives the ΔRMSE of IS errors for OOS period. MSE-F is 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 **all** model where they are obtained from McCracken (2004) (critical values for **ms** model are not calculated). 'Pwr' is the power of ΔRMSE and is calculated as the fraction of draws where the simulated ΔRMSE is greater than the empirically calculated 95% critical value and is reported in percent. The two numbers under the power column are power for all simulations and simulations that are found to be in-sample significant at the 95% level. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

Panel A: Insignificant predictors

	Variable	Data	IS $\overline{R}^2$	OOS ΔRMSE D+20	OOS ΔRMSE 1965–
d/p	Dividend Price Ratio	192701–200412	0.14	0.0014	0.0013
d/e	Dividend Payout Ratio	192701–200412	0.03	−0.0292	−0.0397
svar	Stock Variance	192701–200412	−0.08	−0.0027	−0.0028
tbl	T-Bill Rate	192701–200412	0.12	0.0005*	0.0013
lty	Long Term Yield	192701–200412	−0.00	−0.0189	−0.0204
ltr	Long Term Return	192701–200412	0.03	−0.0192	−0.0059
tms	Term Spread	192701–200412	0.07	0.0039**	0.0065**
dfy	Default Yield Spread	192701–200412	−0.07	−0.0025	0.0010
dfr	Default Return Spread	192701–200412	−0.00	−0.0073	−0.0020
infl	Inflation	192701–200412	−0.00	0.0023*	0.0033

Panel B: Significant predictors

			IS		Forecasts begin 20 years after sample data				Forecasts begin 1965			
					IS for OOS	OOS			IS for OOS	OOS		
	Variable	Data	$\overline{R}^2$	ΔRMSE	ΔRMSE	ΔRMSE	MSE-F	Pwr	ΔRMSE	ΔRMSE	MSE-F	Pwr
d/y	Dividend Yield	192701–200412	0.25*	0.0100*	0.0139**	–0.0059	–1.95	37 (**79**,6)	0.0053	–0.0037	–0.80	33 (**70**,6)
e/p	Earning Price Ratio	192701–200412	0.53**	0.0181**	0.0081	–0.0226	–7.46	58 (**67**,13)	–0.0039	–0.0224	–4.84	55 (**63**,15)
csp	Cross-Sectional Prem	193705–200212	0.92***	0.0246***	0.0128*	–0.0137	–3.45	68 (**85**,12)	0.0201**	0.0215	4.40***	65 (**80**,14)
b/m	Book to Market	192701–200412	0.39**	0.0140**	–0.0054	–0.0327	–10.77	51 (**69**,12)	–0.0172	–0.0485	–10.38	49 (**64**,16)
ntis	Net Equity Expansion	192701–200412	0.71***	0.0231***	0.0066	0.0053	1.76**	64 (**85**,11)	0.0103	–0.0002	–0.04	56 (**74**,12)
all	Kitchen Sink	192701–200412	2.05***	0.0910	0.0644	–0.2348	–72.04	— (—,—)	0.0699	–0.1355	–28.16	— (—,—)
ms	Model Selection	192701–200412	—	—	—	–0.0510	–16.66	— (—,—)	—	–0.0418	–8.97	— (—,—)

Table 4: Forecasts at Monthly Frequency using Campbell and Thompson (2005) procedure

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at the monthly frequency (both in the forecasting equation and forecast) using the procedure of Campbell and Thompson (2005) (henceforth, CT). Variables are explained in Section 2. Stock return is price changes, including dividends, of S&P500. Panel A uses the log returns (as in the rest of the tables) while Panel B uses simple returns (as in CT). The data period is December 1927 to December 2004, except for **csp** (May 1937 to December 2002) and **cay**3 (December 1951 to December 2004). A star next to $\overline{R}^2$ (in percent) denotes significance of the in-sample regression (as measured by empirical F-statistic). Variables are sorted in increasing order of in-sample significance. 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). $\Delta U^{\gamma=3}$ is the utility difference for mean variance utility optimizer with risk aversion coefficient $\gamma = 3$ who trades based on unconditional forecast and conditional forecast. Portfolio weights are denoted by w (a cap $w_{\max} = 150\%$ is imposed on all portfolio weights). ΔRMSE and ΔU are in percent per month while w is in percent. Subscript U is for unconditional forecast, PN is for plain conditional forecast, and CT is the CT conditional forecast. The column titled $\Delta U^{\gamma=x}_{\mathrm{CT}}$ gives the utility based on risk aversion coefficient $\gamma = x$, where x equalizes the U_{mkt} and U_{rf}. The columns titled 'Frcst=' give the fraction of months where the forecast was truncated to zero or set equal to the unconditional mean.The panel header gives the utility of investing based on unconditional forecast (U_0), buy-and hold market (U_{mkt}), and the riskfree asset (U_{rf}).Critical values of all statistics are obtained empirically from bootstrapped distributions, except for **cay**3 model where they are obtained from McCracken (2004). Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

Panel A: Log Return $U_{\mathbf{U}} = 0.6938\%, U_{\mathbf{mkt}} = 0.7009\%, U_{\mathbf{rf}} = 0.3994\%$

	Variable	IS		OOS							
		$\overline{R}^2$	$\Delta U^{\gamma=3}_{\mathrm{PN}}$	$\Delta\mathrm{RMSE}_{\mathrm{PN}}$	$\Delta\mathrm{RMSE}_{\mathrm{CT}}$	$\Delta U^{\gamma=3}_{\mathrm{CT}}$	$\Delta U^{\gamma=7.53}_{\mathrm{CT}}$	Frcst= 0	Frcst= U	$w_{\mathrm{CT}} = w_{\max}$	$\Delta w_{\mathrm{CT}} - \Delta w_{\mathrm{U}}$
svar	Stock Variance	−0.09	0.0228	−0.0037	−0.0029	−0.0192	−0.0104	0.0	53.6	26.1	0.1
dfy	Default Yield Spread	−0.07	0.0088	−0.0046	−0.0033	−0.0286	−0.0177	2.9	25.3	25.5	0.7
lty	Long Term Yield	−0.02	0.0767	−0.0244	0.0045**	0.0647	0.0496	49.2	0.0	8.5	1.3
dfr	Default Return Spread	−0.01	0.0347	−0.0075	−0.0066	−0.0372	−0.0259	1.5	1.8	24.2	22.6
infl	Inflation	−0.00	0.0619	0.0019*	0.0029*	0.0341	0.0286	1.3	0.0	25.8	9.0
ltr	Long Term Return	0.03	0.1332	−0.0191	0.0036*	0.0270	0.0135	5.0	39.7	33.1	25.0
d/e	Dividend Payout Ratio	0.04	−0.0789	−0.0218	−0.0218	−0.0179	−0.1243	0.0	0.0	59.1	−0.2
tbl	T-Bill Rate	0.11	0.1736	−0.0037	0.0025*	0.0699	0.0520	26.1	0.0	5.4	2.8
d/p	Dividend Price Ratio	0.11	0.0579	−0.0006	0.0031	−0.0468	−0.0006	21.2	0.0	4.4	2.7
tms	Term Spread	0.12	0.1793	0.0056**	0.0054**	0.1171	0.0407	3.6	12.0	30.7	5.3
d/y	Dividend Yield	0.22*	0.0406	−0.0091	0.0041*	−0.0659	0.0082	52.1	0.0	9.8	3.2
b/m	Book to Market	0.43**	−0.1312	−0.0434	−0.0266	−0.1762	−0.1654	48.5	0.0	20.6	4.0
e^{10}**/p**	Earning(10Y) Price Ratio	0.45**	−0.0493	−0.0301	−0.0076	−0.1469	−0.0642	55.0	0.0	6.6	3.7
e/p	Earning Price Ratio	0.51**	0.0382	−0.0185	−0.0127	−0.0525	−0.0516	36.9	0.0	27.3	3.1
csp	Cross-Sectional Prem	0.92***	0.1896	−0.0137	0.0101**	0.0986	0.0847	50.2	0.0	9.3	4.1
eqis	Pct Equity Issuing	0.84***	0.1696	0.0097**	0.0128***	0.1575	0.0875	10.2	0.0	37.5	2.1
ntis	Net Equity Expansion	0.89***	0.0086	0.0053**	0.0063**	0.0169	0.0194	5.0	0.0	43.4	5.8
cay3	Cnsmptn, Wlth, Incme	2.73***	0.5159	−0.0023	0.0196**	0.2189	0.1296	37.5	0.0	18.6	10.2

Panel B: Simple Return $U_{\mathrm{U}} = 0.8169\%, U_{\mathbf{mkt}} = 0.7928\%, U_{\mathbf{rf}} = 0.3994\%$

	Variable	IS		OOS							
		$\overline{R}^2$	$\Delta U_{\mathrm{PN}}^{\gamma=3}$	$\Delta\mathrm{RMSE}_{\mathrm{PN}}$	$\Delta\mathrm{RMSE}_{\mathrm{CT}}$	$\Delta U_{\mathrm{CT}}^{\gamma=3}$	$\Delta U_{\mathrm{CT}}^{\gamma=7.53}$	Frcst= 0	Frcst= U	$w_{\mathrm{CT}} = w_{\max}$	$\Delta w_{\mathrm{CT}} - \Delta w_{\mathrm{U}}$
d/e	Dividend Payout Ratio	–0.10	–0.0222	–0.0116	–0.0115	–0.0149	–0.0655	0.0	2.9	59.1	–0.2
svar	Stock Variance	–0.08	–0.0185	–0.0135	–0.0135	–0.0362	–0.0339	0.0	0.0	36.4	2.0
dfr	Default Return Spread	–0.06	0.0386	–0.0050	–0.0036	0.0096	–0.0096	0.0	10.9	46.3	7.6
lty	Long Term Yield	0.03	0.0648	–0.0138	0.0087**	0.0612	0.0777	34.5	0.0	19.7	2.2
ltr	Long Term Return	0.07	0.0945	–0.0095	0.0048**	0.0594	0.0137	3.4	37.5	52.0	24.4
infl	Inflation	0.14	0.0498	0.0037*	0.0049**	0.0422	0.0430	1.3	0.0	44.5	11.3
tms	Term Spread	0.18	0.1577	0.0075**	0.0075**	0.1457	0.0756	3.5	0.0	60.4	4.2
tbl	T-Bill Rate	0.20*	0.1456	0.0016*	0.0082**	0.0964	0.0876	23.1	0.0	16.8	3.1
dfy	Default Yield Spread	0.26*	–0.0080	–0.0091	–0.0077	–0.0739	–0.0310	4.1	0.0	29.9	2.4
d/p	Dividend Price Ratio	0.32*	–0.0100	–0.0033	0.0061*	–0.1026	0.0173	30.4	0.0	16.4	4.1
d/y	Dividend Yield	0.46**	–0.0365	–0.0202	0.0014*	–0.1432	0.0091	53.0	0.0	16.6	3.4
e/p	Earning Price Ratio	0.53**	0.0411	–0.0187	–0.0186	–0.0430	–0.0520	17.8	0.0	35.2	4.3
b/m	Book to Market	0.79***	–0.1671	–0.0646	–0.0442	–0.2252	–0.2058	43.5	0.0	31.8	3.6
eqis	Pct Equity Issuing	0.82***	0.1413	0.0076**	0.0110***	0.1370	0.0873	6.6	0.0	55.0	1.5
e^{10}/p	Earning(10Y) Price Ratio	0.84***	–0.1494	–0.0427	–0.0078	–0.1329	–0.0629	50.5	0.0	15.6	5.0
csp	Cross-Sectional Prem	0.99***	0.1629	–0.0136	0.0079**	0.0631	0.0743	43.6	0.0	13.9	4.9
ntis	Net Equity Expansion	0.94***	0.0143	0.0015*	0.0016*	0.0198	–0.0016	0.4	0.0	57.2	4.6
cay3	Cnsmptn, Wlth, Incme	2.66***	0.5102	–0.0058	0.0168**	0.2008	0.1060	35.5	0.0	24.2	9.7

Table 5: Significant Forecasts Using Various **d/p**, **e/p**, and **d/e** Ratios

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. Stock return is price changes, including dividends, of S&P500 (monthly data uses CRSP data for calculation of stock returns). A star next to $\overline{R}^2$ denotes significance of the in-sample regression (as measured by empirical F-statistic). 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). All ΔRMSE numbers are in percent per frequency corresponding to the column entitled 'Freq'. The 'Freq' column also gives the first year of forecast. MSE-F is the 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. The table reports only those combinations of **d/p** **e/p** and **d/e** that were found to be in-sample significant.

	Variable	Data	Freq	IS $\overline{R}^2$	OOS ΔRMSE	OOS MSE-F
e/p	Earning(1Y) Price Ratio	1927–2004	M 1965–	0.53**	−0.0224	−4.84
e^5/p	Earning(5Y) Price Ratio	1927–2004	M 1965–	0.32*	−0.0086	−1.87
e^{10}/p	Earning(10Y) Price Ratio	1927–2004	M 1965–	0.48**	−0.0135	−2.91
e^3/p	Earning(3Y) Price Ratio	1882–2004	A 1902–	2.46**	−0.0133	−0.14*
e^5/p	Earning(5Y) Price Ratio	1882–2004	A 1902–	2.83**	0.0390	0.42*
e^{10}/p	Earning(10Y) Price Ratio	1882–2004	A 1902–	4.88**	0.2967	3.26**
d^5/p	Dividend(5Y) Price Ratio	1882–2004	A 1902–	2.52*	0.0473	0.51*
d^{10}/p	Dividend(10Y) Price Ratio	1882–2004	A 1902–	2.14*	−0.0042	−0.05*
d/e^{10}	Dividend(1Y) Earning(10Y) Ratio	1882–2004	A 1902–	1.48*	−0.0597	−0.64
e^3/p	Earning(3Y) Price Ratio	1882–2004	A 1965–	2.46**	−0.0861	−0.43
e^5/p	Earning(5Y) Price Ratio	1882–2004	A 1965–	2.83**	−0.2008	−0.99
e^{10}/p	Earning(10Y) Price Ratio	1882–2004	A 1965–	4.88**	−0.6726	−3.19
d^5/p	Dividend(5Y) Price Ratio	1882–2004	A 1965–	2.52*	−0.4591	−2.22
d^{10}/p	Dividend(10Y) Price Ratio	1882–2004	A 1965–	2.14*	−0.4263	−2.07
d/e^{10}	Dividend(1Y) Earning(10Y) Ratio	1882–2004	A 1965–	1.48*	−1.0451	−4.79
e^3/p	Earning(3Y) Price Ratio	1882–2004	5Y 1902–	10.49*	0.7036	3.55**
e^5/p	Earning(5Y) Price Ratio	1882–2004	5Y 1902–	15.32**	0.9691	4.94**
e^{10}/p	Earning(10Y) Price Ratio	1882–2004	5Y 1902–	15.80*	−0.5186	−2.50
d/p	Dividend(1Y) Price Ratio	1882–2004	5Y 1902–	11.83*	−0.0464	−0.23*
d^3/p	Dividend(3Y) Price Ratio	1882–2004	5Y 1902–	12.67*	−0.3145	−1.53*
d^5/p	Dividend(5Y) Price Ratio	1882–2004	5Y 1902–	13.35*	−0.6684	−3.21
e^3/p	Earning(3Y) Price Ratio	1882–2004	5Y 1965–	10.49*	−2.1562	−4.29
e^5/p	Earning(5Y) Price Ratio	1882–2004	5Y 1965–	15.32**	−3.4594	−6.52
e^{10}/p	Earning(10Y) Price Ratio	1882–2004	5Y 1965–	15.80*	−4.0076	−7.39
d/p	Dividend(1Y) Price Ratio	1882–2004	5Y 1965–	11.83*	−4.4347	−8.04
d^3/p	Dividend(3Y) Price Ratio	1882–2004	5Y 1965–	12.67*	−4.2526	−7.77
d^5/p	Dividend(5Y) Price Ratio	1882–2004	5Y 1965–	13.35*	−4.6114	−8.30

Table 6: Forecasts at Monthly Frequency with Alternative Procedures and Total Returns

This table presents statistics on forecast errors in-sample (IS) and out-of-sample (OOS) for excess stock return forecasts at the monthly frequency (both in the forecasting equation and forecast). Variables are explained in Section 2. Stock return is price changes, including dividends, of S&P500 calculated using CRSP data. 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). A star next to $\overline{R}^2$ denotes significance of the in-sample regression (as measured by empirical F-statistic). ΔRMSE is the monthly 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). A star next to ΔRMSE denotes significance based on the MSE-F statistic by McCracken (2004). One-sided critical values of MSE-F statistic are obtained empirically from bootstrapped distributions. 'Pwr' is the power of ΔRMSE and is calculated as the fraction of draws where the simulated ΔRMSE is greater than the empirically calculated 95% critical value and is reported in percent. The two numbers under the power column are power for all simulations and simulations that are found to be in-sample significant at the 95% level. Significance levels at 90%, 95%, and 99% are denoted by one, two, and three stars, respectively.

			OLS				Stambaugh			Lewellen		
			IS		OOS		IS	OOS		IS	OOS	
	Variable	Data	ρ	$\overline{R}^2$	ΔRMSE	Pwr	$\overline{R}^2$	ΔRMSE	Pwr	$\overline{R}^2$	ΔRMSE	Pwr
d/e	Dividend Payout Ratio	192701–200412	0.9985	0.03	−0.0367	15 (**70**,4)	0.03	−0.0385	14 (**69**,4)	0.03	−0.0372	15 (**69**,4)
lty	Long Term Yield	192701–200412	0.9963	−0.00	−0.0209	9 (**69**,3)	−0.01	−0.0338	10 (**68**,4)	−0.01	−0.0184	10 (**69**,4)
d/y	Dividend Yield	192701–200412	0.9930	0.25 *	−0.0061	34 (**70**,6)	0.25 *	−0.0053	33 (**71**,6)	0.25 *	−0.0031	33 (**71**,5)
d/p	Dividend Price Ratio	192701–200412	0.9930	0.14	−0.0008	29 (**55**,11)	0.04	−0.0030	27 (**69**,2)	−0.12 **	−0.0128	5 (**5**,NaN)
tbl	T-Bill Rate	192701–200412	0.9924	0.12	0.0010	19 (**69**,5)	0.11	−0.0025	20 (**69**,6)	0.11	−0.0011	20 (**68**,6)
e/p	Earning Price Ratio	192701–200412	0.9881	0.53 **	−0.0273	55 (**63**,15)	0.47 **	−0.0111	60 (**74**,3)	0.04 ***	0.0039 **	42 (**42**,NaN)
b/m	Book to Market	192701–200412	0.9845	0.39 **	−0.0563	49 (**64**,16)	0.35 *	−0.0369	48 (**71**,3)	−0.14 **	−0.0030	23 (**23**,NaN)
csp	Cross-Sectional Prem	193705–200212	0.9788	0.92 ***	−0.0139	68 (**85**,12)	0.92 ***	−0.0129	68 (**85**,12)	0.92 ***	−0.0094	68 (**85**,10)
dfy	Default Yield Spread	192701–200412	0.9764	−0.07	0.0015	9 (**58**,3)	−0.08	−0.0019	8 (**59**,2)	−0.16	−0.0091	5 (**44**,1)
ntis	Net Equity Expansion	192701–200412	0.9698	0.71 ***	0.0033	56 (**74**,12)	0.71 ***	0.0031	56 (**74**,12)	0.71 ***	0.0022	56 (**72**,14)
tms	Term Spread	192701–200412	0.9569	0.07	0.0075*	21 (**64**,7)	0.07	0.0071*	20 (**64**,6)	0.07	0.0073**	21 (**65**,7)
svar	Stock Variance	192701–200412	0.6002	−0.08	−0.0029	6 (**55**,3)	−0.08	−0.0026	7 (**55**,3)	−1.69 **	−0.0049	7 (**7**,NaN)
infl	Inflation	192701–200412	0.5532	−0.00	0.0030	15 (**61**,5)	−0.00	0.0030	15 (**61**,5)	−0.03	0.0018	16 (**54**,8)
ltr	Long Term Return	192701–200412	0.0551	0.03	−0.0017	17 (**62**,6)	0.03	−0.0016	17 (**62**,6)	−1.61 ***	−0.1483	12 (**12**,NaN)
dfr	Default Return Spread	192701–200412	−0.1994	−0.00	0.0015	14 (**62**,4)	−0.00	0.0015	14 (**62**,4)	−1.27 *	−0.0288	11 (**38**,8)

Table 7: Encompassing Tests

This table presents statistics on encompassing tests for excess stock return forecasts at various frequencies. Variables are explained in Section 2. 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). ΔRMSE* is the RMSE difference between the unconditional forecast and the optimal forecast for the same sample/forecast period. ΔRMSE^{*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.

Panel A: Annual Data

		Estimation: OOS Forecast:	All Data After 20 years					All Data After 1965				After 1927 After 1965				
		Data	$\overline{R}^2$	λ	ENC	ΔRMSE*	ΔRMSE^{*r}	λ	ENC	ΔRMSE*	ΔRMSE^{*r}	$\overline{R}^2$	λ	ENC	ΔRMSE*	ΔRMSE^{*r}
d/p	Dividend Price Ratio	1872–2004	*0.47*	0.20	0.44	0.0074	–0.2604	0.39	0.79	0.0618	–0.5148	*1.60*	0.54	2.03*	0.2188	–0.3740
d/y	Dividend Yield	1872–2004	*0.89*	0.38	1.86	0.0593	–0.5725	0.30	1.14*	0.0699	–0.5542	*2.65**	0.40	3.03**	0.2550	–0.3052
e/p	Earning Price Ratio	1872–2004	*1.00*	0.19	0.33	0.0054	–0.2305	0.63	1.08*	0.1346	–0.5094	*3.01**	0.46	2.27**	0.2123	–0.4387
d/e	Dividend Payout Ratio	1872–2004	*–0.75*	–1.75	–1.44	0.2169	0.0985	–9.42	–0.44	0.8540	0.4304	*–1.24*	–5.00	–1.21	1.3658	0.8920
svar	Stock Variance	1885–2004	*–0.76*	–0.42	–4.69	0.2405	–0.6465	2.06	0.03	0.0141	–0.6153	*–1.33*	–16.60	–0.17	0.5883	0.4513
b/m	Book to Market	1921–2004	*3.01**	0.48	3.88**	0.2360	–0.0793	0.18	1.05	0.0424	–0.8211	*3.97**	0.17	1.44*	0.0565	–0.5075
ntis	Net Equity Expansion	1927–2004	*–0.03*	0.39	0.47	0.0260	–0.2349	–0.45	–0.27	0.0248	–0.6602	*–0.03*	–0.45	–0.27	0.0248	–0.6602
eqis	Pct Equity Issuing	1927–2004	*9.62***	0.72	4.60**	0.4430	–0.0109	0.61	3.27**	0.3938	–0.6814	*9.62****	0.61	3.27**	0.3938	–0.6814
tbl	T-Bill Rate	1920–2004	*0.57*	0.42	2.36*	0.1252	–1.2528	0.44	2.36**	0.2157	–1.2916	*0.37*	0.35	2.92**	0.2185	–0.5465
lty	Long Term Yield	1919–2004	*–0.53*	0.30	2.86*	0.1117	–0.7032	0.30	2.54**	0.1676	–0.9343	*–0.86*	0.26	2.53**	0.1517	–0.5650
ltr	Long Term Return	1926–2004	*1.00*	0.31	4.50**	0.2112	–0.1444	0.24	2.41**	0.1334	–8.5612	*0.93*	0.25	2.40**	0.1383	–8.8608
tms	Term Spread	1920–2004	*0.30*	0.42	1.03	0.0536	–1.0314	0.50	1.15*	0.1174	–0.8751	*1.06*	0.52	2.05**	0.2148	–0.5269
dfy	Default Yield Spread	1919–2004	*–1.20*	–2.65	–0.47	0.1550	–0.9710	–10.78	–0.29	0.6595	0.5154	*–1.33*	–12.10	–0.23	0.5851	0.4633
dfr	Default Return Spread	1926–2004	*0.38*	0.44	0.85	0.0506	–0.3715	0.47	0.75	0.0719	–0.3862	*0.30*	0.48	0.72	0.0700	–0.3927
infl	Inflation	1919–2004	*–0.98*	–2.35	–0.65	0.1897	–0.4707	–1.24	–0.13	0.0315	–15.3729	*–1.05*	–3.34	–0.87	0.6255	–0.4060
all	Kitchen Sink	1927–2004	*15.61***	0.14	5.02	0.1753	0.0051	–0.01	–0.20	0.0008	–2.0894	*15.61***	–0.01	–0.20	0.0008	–2.0894
caya	Cnsmptn, Wlth, Incme	1948–2001	—	0.43	6.14***	0.6810	0.0028	0.43	6.14***	0.6810	0.0028	—	0.43	6.14***	0.6810	0.0028
ms	Model Selection	1927–2004	—	0.40	9.03	0.5511	0.3113	0.30	3.70	0.2559	–0.2321	—	0.30	3.70	0.2559	–0.2321
cayp	Cnsmptn, Wlth, Incme	1948–2001	*24.89***	0.82	15.27***	2.3716	1.6803	0.82	15.27***	2.3716	1.6803	*24.89****	0.82	15.27***	2.3716	1.6803

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	192701–200412	*0.14*	0.53	3.88**	0.0062	−0.0139	0.53	2.51**	0.0061	−0.0114
d/y	Dividend Yield	192701–200412	*0.25**	0.43	6.20***	0.0081	−0.0119	0.45	3.71**	0.0077	−0.0087
e/p	Earning Price Ratio	192701–200412	*0.53***	0.35	9.09***	0.0098	−0.0136	0.27	2.94**	0.0038	−0.0177
d/e	Dividend Payout Ratio	192701–200412	*0.03*	−0.01	−0.06	0.0000	−0.0148	−1.12	−2.94	0.0154	0.0003
svar	Stock Variance	192701–200412	*−0.08*	−11.74	−0.43	0.0152	0.0021	−12.62	−0.29	0.0170	0.0047
csp	Cross-Sectional Prem	193705–200212	*0.92****	0.39	6.33***	0.0099	−0.0133	0.83	5.55***	0.0225	−0.0004
b/m	Book to Market	192701–200412	*0.39***	0.17	2.79*	0.0015	−0.0420	0.06	0.75	0.0002	−0.0265
ntis	Net Equity Expansion	192701–200412	*0.71****	0.63	4.28**	0.0081	−0.0055	0.50	2.77**	0.0063	−0.0179
tbl	T-Bill Rate	192701–200412	*0.12*	0.51	5.38**	0.0082	−0.0231	0.52	4.81***	0.0114	−0.0228
lty	Long Term Yield	192701–200412	*−0.00*	0.35	7.56***	0.0081	−0.0085	0.36	5.50***	0.0091	−0.0162
ltr	Long Term Return	192701–200412	*0.03*	−0.12	−0.59	0.0002	−0.0134	0.31	1.08*	0.0016	−0.0246
tms	Term Spread	192701–200412	*0.07*	0.69	2.39**	0.0049	−0.0307	0.73	2.23**	0.0075	−0.0411
dfy	Default Yield Spread	192701–200412	*−0.07*	−1.21	−0.30	0.0011	−0.0068	1.76	0.15	0.0012	−0.0186
dfr	Default Return Spread	192701–200412	*−0.00*	−0.63	−0.68	0.0013	−0.0141	0.10	0.05	0.0000	−0.0215
infl	Inflation	192701–200412	*−0.00*	1.11	0.71	0.0024	−0.0116	1.31	0.59	0.0035	−0.0552
all	Kitchen Sink	192701–200412	*2.05****	0.07	5.76*	0.0013	−0.0152	0.16	6.58**	0.0051	−0.0316
ms	Model Selection	192701–200412	—	0.05	0.96	0.0002	−0.0237	0.11	1.27	0.0007	−0.0271

References

Ang, Andrew, and Geert Bekaert, 2003, Stock Return Predictability: Is it There?, Working paper, Columbia Business School.

Avramov, Doron, 2002, Stock Return Predictability and Model Uncertainty, Journal of Financial Economics 64, 423–458.

Baker, Malcolm, Ryan Taliaferro, and Jeffrey Wurgler, 2004, Pseudo Market Timing and Predictive Regressions, Working paper, Harvard Business School and New York University.

Baker, Malcom, and Jeffrey Wurgler, 2000, The Equity Share in New Issues and Aggregate Stock Returns, Journal of Finance 55, 2219–2257.

Ball, Ray, 1978, Anomalies in Relationship Between Securities' Yields and Yield-Surrogates, Journal of Financial Economics 6, 103–126.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, Review of Financial Studies 12, 405–428.

Boudoukh, Jacob, Roni Michaely, Matthew P. Richardson, and Michael R. Roberts, 2003, On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing, Working paper, Arison School of Business, IDC and Cornell University and New York University and Universitly of Pennsylvania.

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, Economic Significance of Predictable Variations in Stock Index Returns, Journal of Finance 64, 1177–1189.

Brennan, Michael J., and Yihong Xia, 2004, Persistence, Predictability, and Portfolio Planning, Working Paper, UCLA and Wharton.

Butler, Alexander W., Gustavo Grullon, and James Weston, 2004, Can Managers Forecast Aggregate Market Returns, Working Paper, forthcoming in Journal of Finance.

Campbell, John Y., 1987, Stock returns and the Term Structure, Journal of Financial Economics 18, 373–399.

Campbell, John Y., and Robert J. Shiller, 1988a, Stock Prices, Earnings, and Expected Dividends, Journal of Finance 43, 661–676.

Campbell, John Y., and Robert J. Shiller, 1988b, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, Review of Financial Studies 1, 195–227.

Campbell, John Y., and Robert J. Shiller, 1998, Valuation Ratios and the Long-Run Stock Market Outlook, Journal of Portfolio Management 24, 11–26.

Campbell, John Y., and Samuel B. Thompson, 2005, Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?, Working paper, Harvard University.

Campbell, John Y., and Luis M. Viceira, 2002, Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. (Oxford University Press Oxford, UK).

Campbell, John Y., and Tuomo Vuolteenaho, 2004, Inflation Illusion and Stock Prices, American Economic Review 94, 19–23.

Campbell, John Y., and Motohiro Yogo, 2003, Efficient Tests of Stock Return Predictability, Working paper, Harvard University.

Clark, Todd E., and Michael W. McCracken, 2001, Tests of Forecast Accuracy and Encompassing for Nested Models, Journal of Econometrics 105, 85–110.

Cochrane, John H., 1997, Where is the Market Going? Uncertain Facts and Novel Theories, Federal Reserve Bank of Chicago - Economic Perspectives 21, 3–37.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing Predictive Accuracy, Journal of Business & Economic Statistics 13, 253–263.

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

Fama, Eugene F., 1981, Stock Returns, Real Activity, Inflation, and Money, American Economic Review 71, 545–565.

Fama, Eugene F., and Kenneth R. French, 1988, Dividend Yields and Expected Stock Returns, Journal of Financial Economics 22, 3–25.

Fama, Eugene F., and Kenneth R. French, 1989, Business Conditions and Expected Returns on Stocks and Bonds, Journal of Financial Economics 25, 23–49.

Fama, Eugene F., and G. William Schwert, 1977, Asset Returns and Inflation, Journal of Financial Economics 5, 115–146.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious Regressions in Financial Economics, The Journal of Finance 58, 1393–1413.

Goetzmann, William N., and Philippe Jorion, 1993, Testing the Predictive Power of Dividend Yields, Journal of Finance 48, 663–679.

Goyal, Amit, and Ivo Welch, 2003, Predicting the Equity Premium with Dividend Ratios, Management Science 49, 639–654.

Guo, Hui, 2006, On the Out-of-Sample Predictability of Stock Market Returns, Journal of Business 79, forthcoming.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold, 1997, Testing the Equality of Prediction Mean Squared Errors, International Journal of Forecasting 13, 281–291.

Hodrick, Robert J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies 5, 257–286.

Keim, Donald, and Ananth Madhavan, 1997, Transaction costs and investment style: an inter-exchange analysis of institutional equity trades, Journal of Financial Economics 46, 265–292.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets, Journal of Financial Economics 17, 357–390.

Kilian, Lutz, 1999, Exchange Rates and Monetary Fundamentals: What Do We Learn From LongHorizon Regressions?, Journal of Applied Econometrics 14, 491–510.

Kothari, S.P., and Jay Shanken, 1997, Book-to-market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis, Journal of Financial Economics 44, 169–203.

Lamont, Owen, 1998, Earnings and Expected Returns, Journal of Finance 53, 1563–1587.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth, and Expected Stock Returns, Journal of Finance 56, 815–849.

Lewellen, Jonathan, 2004, Predicting Returns with Financial Ratios, Journal of Financial Economics 74, 209–235.

Lewellen, Jonathan, and Jay Shanken, 2002, Learning, Asset-Pricing Tests, and Market Efficiency, Journal of Finance 57, 1113–1145.

Lintner, John, 1975, Inflation and Security Returns, Journal of Finance 30, 259–280.

Mark, Nelson C., 1995, Exchange Rates and Fundamentals: Evidence on LongHorizon Predictability, American Economic Review 85, 201–218.

McCracken, Michael W., 2004, Asymptotics for Out-of-Sample Tests of Causality, Working paper, University of Missouri-Columbia.

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding Predictability, Journal of Political Economy 112, 1–47.

Nelson, Charles R., and Myung J. Kim, 1993, Predictable Stock Returns: The Role of Small Sample Bias, Journal of Finance 48, 641–661.

Pesaran, Hashem M., and Allan Timmerman, 1995, Predictability of Stock Returns: Robustness and Economic Significance, Journal of Finance 50, 1201–1228.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2006, Cross-Sectional Forecasts of the Equity Premium, Journal of Financial Economics ?, ?

Pontiff, Jeffrey, and Lawrence D. Schall, 1998, Book-to-Market Ratios as Predictors of Market Returns, Journal of Financial Economics 49, 141–160.

Rapach, David E., and Mark E. Wohar, 2004, In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining, Working Paper, forthcoming in Journal of Empirical Finance.

Rissanen, Jorma, 1986, Order Estimation by Accumulated Prediction Errors, Journal of Applied Probability 23A, 55–61.

Rozeff, Michael S., 1984, Dividend Yields Are Equity Risk Premiums, Journal of Portfolio Management 11, 68–75.

Shiller, Robert J., 1984, Stock Prices and Social Dynamics, Brookings Papers on Economic Activity 2, 457–498.

Stambaugh, Robert F., 1999, Predictive Regressions, Journal of Financial Economics 54, 375–421.

Torous, Walter, and Rossen Valkanov, 2000, Boundaries of Predictability: Noisy Predictive Regressions, Working paper, UCLA.

Valkanov, Rossen, 2003, Long-Horizon Regressions: Theoretical Results and Applications, JFE 68, 201–232.

Xia, Yihong, 2001, Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation, Journal of Finance 56, 205–246.