Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy

David E. Rapach
Saint Louis University
rapachde@slu.edu

Jack K. Strauss
Saint Louis University
jkstraus@slu.edu

Guofu Zhou*
Washington University in St. Louis
zhou@wustl.edu

April 8, 2009

*Corresponding author. Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We are grateful to Andrew Ang, John Campbell, Long Chen, Ethan Chiang, John Cochrane, Hui Guo, Michael Halling, Yufeng Han, Campbell Harvey, Huiyu Huang, Ravi Jagannathan, Augio Kong, Sydney Ludvigson, Raymond Kan, Mike McCracken, Ľuboš Pástor, Jun Tu, seminar participants at McGill University, Singapore Management University, Tsinghua University, Washington University, and the Missouri Economics Conference for helpful comments. We are also especially grateful to an anonymous referee for valuable comments that helped to substantially improve the paper, as well as Matthew Spiegel (the Executive Editor) for insightful comments that led to the paper's investigation of links to the real economy. Rapach and Strauss acknowledge financial support from the Simon Center for Regional Forecasting at Saint Louis University. This paper supersedes an earlier version that circulated under the title, "Out-of-Sample Equity Premium Prediction: Consistently Beating the Historical Average," originally dated July 23, 2007.

Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy

Abstract

Welch and Goyal (2008) find that numerous economic variables with in-sample predictive ability for the equity premium fail to deliver consistent out-of-sample forecasting gains relative to the historical average. Arguing that model uncertainty and instability seriously impair the forecasting ability of individual predictive regression models, we recommend *combining* individual forecasts. Combining delivers statistically and economically significant out-of-sample gains relative to the historical average consistently over time. We provide two empirical explanations for the benefits of forecast combination: (i) combining forecasts incorporates information from numerous economic variables while substantially reducing forecast volatility; (ii) combination forecasts are linked to the real economy.

JEL classifications: C22, C53, G11, G12

Keywords: Out-of-sample equity premium predictability; Predictive regression model; Combination forecasts; Business cycle; Structural breaks

Forecasting stock returns is of great interest to both academics and practitioners in finance, and numerous economic variables have been proposed as predictors of stock returns in the literature. Examples include valuation ratios, such as the dividend-price [Dow (1920), Fama and French (1988, 1989)], earnings-price [Campbell and Shiller (1988, 1998)], and book-to-market [Kothari and Shanken (1997), Pontiff and Schall (1998)], as well as nominal interest rates [Fama and Schwert (1977), Campbell (1987), Breen, Glosten, and Jagannathan (1989), Ang and Bekaert (2007)], the inflation rate [Nelson (1976), Fama and Schwert (1977), Campbell and Vuolteenaho (2004)], term and default spreads [Campbell (1987), Fama and French (1989)], corporate issuing activity [Baker and Wurgler (2000), Boudoukh, Michaely, Richardson, and Roberts (2007)], consumption-wealth ratio [Lettau and Ludvigson (2001)], and stock market volatility [Guo (2006)].[1] Most existing studies focus on in-sample tests and conclude that there is significant evidence of return predictability.

Return predictability remains controversial, however, as emphasized by Spiegel (2008) in a review of recent articles on the topic in the *Review of Financial Studies*.[2] Among these studies, Welch and Goyal (2008) show that a long list of predictors from the literature are unable to deliver consistently superior *out-of-sample* forecasts of the U.S. equity premium relative to a simple forecast based on the historical average (constant expected equity premium model). Their comprehensive study forcefully echoes the typically negative findings of the relatively few studies that consider out-of-sample tests of return predictability. For example, Bossaerts and Hillion (1999) fail to find significant evidence of out-of-sample predictive ability in a collection of industrialized countries for a number of variables for 1990–1995, and Goyal and Welch (2003) find that the dividend-price ratio is not a robust out-of-sample predictor of the U.S. equity premium.[3] The lack of consistent out-of-sample evidence in Welch and Goyal (2008) indicates the need for improved forecasting methods to better establish the empirical reliability of equity premium predictability.

In this paper, we propose a *combination* approach to the out-of-sample equity premium forecasting problem, and explore both its econometric underpinnings and macroeconomic links. To see the intuition behind forecast combination, consider two predictive regression model forecasts, one based on the dividend yield and the other on the term spread. Fama and French (1989) and others show that these variables can detect changes in economic conditions that potentially signal fluctuations in the equity risk premium. But the dividend yield or term spread alone could capture different components of business conditions, and a given individual economic variable may give a number of "false signals" and/or imply an implausible equity risk premium during certain periods. If individual forecasts based on the dividend yield and term spread are weakly correlated, an average of the two forecasts—a simple type of forecast combination—should be less volatile and more reliably track movements in the equity risk premium. This argument can be extended to the many individual economic variables considered as equity premium predictors in the literature.

In general, numerous factors—including many economic variables with potential predictive information, as well as structural instabilities resulting from institutional change, policy shocks, advances in information technology, and investor learning—give rise to a highly uncertain, complex, and constantly evolving data-generating process for expected equity returns that is difficult to approximate with a single predictive regression model.[4] In such an uncertain and unstable environment, while reliance on a single model may yield reasonable forecasts during particular periods, it is unlikely to generate reliable forecasts over time. Along this line, Welch and Goyal (2008) attribute the inconsistent out-of-sample performance of individual predictive regression models to structural instability. We contend that combining across individual forecasts provides a solution that reduces the uncertainty/instability risk associated with reliance on a single model.[5] Indeed, we show that various combinations of forecasts from 15 individual predictive regression models, each based on an economic variable from the literature, generate consistent and significant out-of-sample gains relative to the historical average. This is true using both statistical and economic criteria and holds across a number of historical periods, including more recent periods when the out-of-sample predictive ability of many individual variables is particularly poor. By combining individual predictive regression model forecasts, we thus find that economic variables collectively are valuable and consistently outperform the historical average forecast of the equity premium.

We employ forecast encompassing tests to elucidate the econometric sources of the benefits of forecast combination. These tests produce evidence of significant information differences across individual predictive regression models, so combining individual forecasts improves information content. In addition, we demonstrate that forecast combination stabilizes individual predictive regression model forecasts of the equity premium, much like diversification across individual assets reduces a portfolio's variance, and this lowers the forecast variance relative to any of the individual predictive regression model forecasts. At the same time, the combination forecast has a smaller bias than almost all of the individual forecasts. From a mean square prediction error (MSPE) perspective, both the reduction in forecast variance and relatively small forecast bias enable the combination forecast to significantly outperform the historical average benchmark on a consistent basis over time.

Furthermore, we show that combination forecasts of the equity premium are linked to the real economy, thus providing insights on the economic sources of equity premium predictability. We examine the links in several ways. First, Fama and French (1989) and Cochrane (1999, 2007) contend that heightened risk aversion during economic downturns requires a higher risk premium, thereby generating equity premium predictability. In line with this, we argue that equity risk premium forecasts based on the combination approach are very plausible, with distinct local maxima (minima) of the combination forecasts occurring very near NBER-dated business-cycle troughs (peaks). Relative to combination forecasts, individual predictive regression models produce equity risk premium forecasts with implausibly large fluctuations, while the historical average produces a forecast that is too "smooth," thereby ignoring fluctuations in the risk premium corresponding to business-cycle fluctuations. Combination forecasts of the equity premium are also significantly correlated with future growth in a number of macroeconomic variables, including real GDP, real profits, and real net cash flows.

Second, in the spirit of Liew and Vassalou (2000), we use the above three macroeconomic variables to define "good," "normal," and "bad" growth periods. We find that out-of-sample gains corresponding to combination forecasts of the equity premium are especially evident during bad growth periods, again tying combination forecasts to business-cycle fluctuations.

Third, as stressed by Cochrane (2007), equity premium forecasts are more plausibly related to macroeconomic risk if equity premium predictors can also forecast business cycles. We demonstrate that the same set of 15 economic variables used to form combination forecasts of the equity premium also generate consistent significant out-of-sample gains when forming combination forecasts of real GDP, real profit, and real net cash flow growth. This directly ties equity premium forecasts to forecasts of the real economy based on the same economic variables and same combination approach, suggesting that the usefulness of combination forecasts with respect to equity premium prediction stems in part from their ability to forecast the real economy.

Finally, as argued above, instabilities in predictive regression models of the equity premium help to explain the advantages of combination forecasts. Interestingly, we show that instabilities in individual equity premium predictive regression models are related to instabilities in the real economy. More specifically, using the Bai and Perron (1998) methodology, we find extensive evidence of structural breaks in individual predictive regression models of real GDP, real profit, and real net cash flow growth based on the same set of 15 economic variables used to predict the equity premium. Moreover, these structural breaks are frequently significant in individual predictive regression models of the equity premium. Overall, links between combination forecasts of the equity premium and the real economy provide additional support for the combination approach.

The remainder of the paper is organized as follows. Section 1 outlines the econometric methodology. The out-of-sample forecasting results for the individual predictive regression models and combining methods are reported in Section 2. Section 3 shows how combination forecasts incorporate useful information from multiple economic variables while reducing forecast volatility. Section 4 examines links between combination forecasts of the equity premium and the real economy. Section 5 concludes.

1 Econometric Methodology

In this section, we discuss the predictive regression model framework, then forecast combination, including a short review of the literature, and, finally, the criteria we use to evaluate the out-of-sample forecasts.

1.1 Predictive Regression Model

We begin with a standard predictive regression model for the equity premium, which can be expressed as

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the return on a stock market index in excess of the risk-free interest rate, $x_{i,t}$ is a variable whose predictive ability is of interest, and ε_{t+1} is a disturbance term. As in Welch and Goyal (2008), we generate out-of-sample forecasts of the equity premium using a recursive (expanding) estimation window. More specifically, we first divide the total sample of T observations for r_t and $x_{i,t}$ into an in-sample portion composed of the first m observations and an out-of-sample portion composed of the last q observations. The initial out-of-sample forecast of the equity premium based on the predictor $x_{i,t}$ is given by

$$\hat{r}_{i,m+1} = \hat{\alpha}_{i,m} + \hat{\beta}_{i,m} x_{i,m}, \tag{2}$$

where $\hat{\alpha}_{i,m}$ and $\hat{\beta}_{i,m}$ are the ordinary least squares (OLS) estimates of α_i and β_i, respectively, in (1) generated by regressing $\{r_t\}_{t=2}^{m}$ on a constant and $\{x_{i,t}\}_{t=1}^{m-1}$. The next out-of-sample forecast is given by

$$\hat{r}_{i,m+2} = \hat{\alpha}_{i,m+1} + \hat{\beta}_{i,m+1} x_{i,m+1}, \tag{3}$$

where $\hat{\alpha}_{i,m+1}$ and $\hat{\beta}_{i,m+1}$ are generated by regressing $\{r_t\}_{t=2}^{m+1}$ on a constant and $\{x_{i,t}\}_{t=1}^{m}$. Proceeding in this manner through the end of the out-of-sample period, we generate a series of q out-of-sample forecasts of the equity premium based on $x_{i,t}$, $\{\hat{r}_{i,t+1}\}_{t=m}^{T-1}$. We emphasize that this out-of-sample forecasting exercise simulates the situation of a forecaster in real time. In our empirical applications in Section 2 below, we generate out-of-sample forecasts of the equity premium using 15 individual predictive regression models ($i = 1, \ldots, N$ and $N = 15$), where each model is based on one of the 15 variables from Welch and Goyal (2008) for which quarterly data are available for 1947:1–2005:4.[6]

Following Welch and Goyal (2008) and Campbell and Thompson (2008), the historical average of the equity premium, $\bar{r}_{t+1} = \sum_{j=1}^{t} r_j$, serves as a natural benchmark forecasting model corresponding to a constant expected equity premium. Intuitively, if $x_{i,t}$ contains information useful for predicting the equity premium, then $\hat{r}_{i,t+1}$ should perform better than $\bar{r}_{t+1}$. Measures to compare $\hat{r}_{i,t+1}$ to $\bar{r}_{t+1}$ are provided in Section 1.3 below.

1.2 Forecast Combination

We utilize information across individual forecasts via forecast combining methods. As pointed out in the seminal paper by Bates and Granger (1969), combinations of individual forecasts can outperform the individual forecasts themselves. Forecast combination has recently received renewed attention in the macroeconomic forecasting literature; see, for example, Stock and Watson (1999, 2003, 2004) with respect to forecasting inflation and real output growth.[7] Despite the increasing popularity of forecast combination in economics, applications in the finance literature are relatively rare. An important example is Mamaysky, Spiegel, and Zhang (2007), who find that combining predictions from an OLS model and the Kalman filter model of Mamaysky, Spiegel, and Zhang (2008) significantly increases the number of mutual funds with predictable out-of-sample alphas. In contrast, our paper focuses on the use of forecast combination to improve equity premium forecasts and examines both its econometric underpinnings and macroeconomic links.[8]

The combination forecasts of r_{t+1} made at time t are weighted averages of the N individual forecasts based on (1):

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \hat{r}_{i,t+1}, \tag{4}$$

where $\{\omega_{i,t}\}_{i=1}^N$ are the *ex ante* combining weights formed at time t. Some of the combining methods require a holdout period to estimate the combining weights, and we use the first q_0 observations from the out-of-sample period as the initial holdout period. For each of the combining methods, we compute combination forecasts over the post-holdout out-of-sample period, leaving us with a total of $q - q_0$ combination forecasts available for evaluation. With one exception (the mean combination forecast described below), all of the combination forecasts allow the combining weights to change at each t. As we discuss below, however, it is typically desirable to have relatively stable combining weights over time.

The combining methods we consider differ in how the weights are determined and can be organized into two classes. The first class uses simple averaging schemes: mean, median, and trimmed mean. The mean combination forecast sets $\omega_{i,t} = 1/N$ for $i = 1, \ldots, N$ in (4), the median combination forecast is the median of $\{\hat{r}_{i,t+1}\}_{i=1}^N$, and the trimmed mean combination forecast sets $\omega_{i,t} = 0$ for the individual forecasts with the smallest and largest values and $\omega_{i,t} = 1/(N-2)$ for the remaining individual forecasts in (4).

The second class of combining methods is based on Stock and Watson (2004), where the combining weights formed at time t are functions of the historical forecasting performance of the individual models over the holdout out-of-sample period. Their discount mean square prediction error (DMSPE) combining method employs the following weights:

$$\omega_{i,t} = \phi_{i,t}^{-1} / \sum_{j=1}^{N} \phi_{j,t}^{-1}, \tag{5}$$

where

$$\phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2, \tag{6}$$

and θ is a discount factor. The DMSPE method thus assigns greater weights to individual predictive regression model forecasts that have lower MSPE values (better forecasting performance) over the holdout out-of-sample period. When $\theta = 1$, there is no discounting, and (5) produces the optimal combination forecast derived by Bates and Granger (1969) for the case where the individual forecasts are uncorrelated. When $\theta < 1$, greater weight is attached to the recent forecast accuracy of the individual models. We consider the two values of 1.0 and 0.9 for θ.

We focus on these two classes of combining methods. Though we considered other methods, where the combining weights are selected more elaborately using in-sample model fit, they performed poorly compared to the simpler schemes. This agrees with the forecasting literature, which indicates that simple combining methods typically outperform more complicated methods [Timmermann (2006)]. For example, we considered a combining method where the weights are functions of the Schwarz information criterion computed for each individual prediction regression model over the estimation period, which is tantamount to setting the combining weights to the approximate in-sample posterior model probabilities [Draper (1995)]. The combining weights for this approach were highly unstable over time, likely reflecting an over-reliance on in-sample fit in the presence of structural instability.[9]

1.3 Forecast Evaluation

We use the out-of-sample R^2 statistic, R^2_{OS}, suggested by Campbell and Thompson (2008) to compare the $\hat{r}_{t+1}$ and $\bar{r}_{t+1}$ forecasts, where $\hat{r}_{t+1}$ is either an individual forecast based on the predictive regression model (1) or a combination forecast. The R^2_{OS} statistic is akin to the familiar in-sample R^2 statistic and is given by

$$R^2_{OS} = 1 - \frac{\sum_{k=q_0+1}^{q}(r_{m+k} - \hat{r}_{m+k})^2}{\sum_{k=q_0+1}^{q}(r_{m+k} - \bar{r}_{m+k})^2}. \tag{7}$$

The R^2_{OS} statistic measures the reduction in MSPE for the predictive regression model or combination forecast relative to the historical average forecast. Thus, when $R^2_{OS} > 0$, the $\hat{r}_{t+1}$ forecast outperforms the historical average forecast according to the MSPE metric.

We further test whether the predictive regression model or combination forecast has a significantly lower MSPE than the historical average benchmark forecast, which is tantamount to testing the null hypothesis that $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$. The most popular method is the Diebold and Mariano (1995) and West (1996) statistic, which has an asymptotic standard normal distribution when comparing forecasts from non-nested models. Clark and McCracken (2001) and McCracken (2007), however, show that this statistic has a non-standard distribution when comparing forecasts from *nested* models, as is clearly the case when comparing predictive regression model forecasts of the equity premium to the historical average: setting $\beta_i = 0$ in (1) yields a model with a constant expected equity premium. Clark and West (2007) develop an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic—what they label the *MSPE-adjusted* statistic—that in conjunction with the standard normal distribution generates asymptotically valid inferences when comparing forecasts from nested linear models.[10] The *MSPE-adjusted* statistic is conveniently calculated by first defining

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2]. \tag{8}$$

By regressing $\{f_{s+1}\}_{s=m+q_0}^{T-1}$ on a constant, and calculating the t-statistic corresponding to the constant, a p-value for a one-sided (upper-tail) test is obtained with the standard normal distribution. In Monte Carlo simulations, Clark and West (2007) demonstrate that the *MSPE-adjusted* statistic performs reasonably well in terms of size and power when comparing forecasts from nested linear models for a variety of sample sizes.

Even if there is evidence that R^2_{OS} is significantly greater than zero, its values are typically small for predictive regression models. This raises the issue of economic significance. Campbell and Thompson (2008) argue that even very small positive R^2_{OS} values, such as 0.5% for monthly data and 1% for quarterly data, can signal an economically meaningful degree of return predictability in terms of increased annual portfolio returns for a mean-variance investor. This provides a simple assessment of forecastability in practice.

A limitation to the R^2_{OS} measure is that it does not explicitly account for the risk borne by an investor over the out-of-sample period. To address this, following Marquering and Verbeek (2004), Campbell and Thompson (2008), Welch and Goyal (2008), and Wachter and Warusawitharana

(2009), we also calculate realized utility gains for a mean-variance investor on a real-time basis. More specifically, we first compute the average utility for a mean-variance investor with relative risk aversion parameter γ who allocates her portfolio monthly between stocks and risk-free bills using forecasts of the equity premium based on the historical average. This exercise requires the investor to forecast the variance of stock returns, and similar to Campbell and Thompson (2008), we assume that the investor estimates the variance using a ten-year rolling window of quarterly returns. A mean-variance investor who forecasts the equity premium using the historical average will decide at the end of period t to allocate the following share of her portfolio to equities in period $t+1$:

$$w_{0,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\bar{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{9}$$

where $\hat{\sigma}^2_{t+1}$ is the rolling-window estimate of the variance of stock returns.[11] Over the out-of-sample period, the investor realizes an average utility level of

$$\hat{\nu}_0 = \hat{\mu}_0 - \left(\frac{1}{2}\right)\gamma\hat{\sigma}^2_0, \tag{10}$$

where $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the benchmark portfolio formed using forecasts of the equity premium based on the historical average.

We then compute the average utility for the same investor when she forecasts the equity premium using an individual predictive regression model or combining method. She will choose an equity share of

$$w_{j,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{11}$$

and realizes an average utility level of

$$\hat{\nu}_j = \hat{\mu}_j - \left(\frac{1}{2}\right)\gamma\hat{\sigma}^2_j, \tag{12}$$

where $\hat{\mu}_j$ and $\hat{\sigma}^2_j$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the portfolio formed using forecasts of the equity premium based on an individual predictive regression model or combining method indexed by j.

In our applications below, we measure the utility gain as the difference between (12) and (10), and we multiply this difference by 400 to express it in average annualized percentage return. The utility gain (or certainty equivalent return) can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive regression model or combination forecast relative to the information in the historical equity premium alone. We report results for $\gamma = 3$; the results are qualitatively similar for other reasonable γ values.

2 Empirical Results

This section describes the data and presents the out-of-sample results for individual predictive regression model and combination forecasts.

2.1 Data

The quarterly data are from Welch and Goyal (2008), who provide detailed descriptions of the data and their sources.[12] Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends, and the Treasury bill rate is used to compute the equity premium. With respect to the economic variables used to predict the equity premium, we consider the 15 variables from Welch and Goyal (2008) for which quarterly data are available for 1947:1–2005:4.

- *Dividend-price ratio (log)*, D/P: difference between the log of dividends paid on the S&P 500 index and log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum.

- *Dividend yield (log)*, D/Y: difference between the log of dividends and log of lagged stock prices.

- *Earnings-price ratio (log)*, E/P: difference between the log of earnings on the S&P 500 index and log of stock prices, where earnings are measured using a one-year moving sum.
- *Dividend-payout ratio (log)*, D/E: difference between the log of dividends and log of earnings.
- *Stock variance*, $SVAR$: sum of squared daily returns on the S&P 500 index.
- *Book-to-market ratio*, B/M: ratio of book value to market value for the Dow Jones Industrial Average.
- *Net equity expansion*, $NTIS$: ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
- *Treasury bill rate*, TBL: interest rate on a 3-month Treasury bill (secondary market).
- *Long-term yield*, LTY: long-term government bond yield.
- *Long-term return*, LTR: return on long-term government bonds.
- *Term spread*, TMS: difference between the long-term yield and Treasury bill rate.
- *Default yield spread*, DFY: difference between BAA- and AAA-rated corporate bond yields.
- *Default return spread*, DFR: difference between long-term corporate bond and long-term government bond returns.
- *Inflation*, $INFL$: calculated from the CPI (all urban consumers); following Welch and Goyal (2008), since inflation rate data are released in the following month, we use $x_{i,t-1}$ in (1) for inflation.
- *Investment-to-capital ratio*, I/K: ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the entire economy [Cochrane (1991)].

We consider three different out-of-sample forecast evaluation periods. Two of the periods correspond to those analyzed in Welch and Goyal (2008): (i) a "long" out-of-sample period covering 1965:1–2005:4; (ii) a more recent out-of-sample period covering the last 30 years of the full sample, 1976:1–2005:4. Welch and Goyal's motivation for considering this latter period is their finding that the out-of-sample predictive ability of a number of the economic variables deteriorates markedly after the Oil Shock of the mid 1970s. With this in mind, we also evaluate a very recent out-of-sample period covering the last six years of the full sample, 2000:1–2005:4, that allows us to analyze how the predictors fare over the recent market period characterized by the collapse of the "technology bubble." Overall, the consideration of multiple out-of-sample periods helps to provide us with a good sense of the robustness of the out-of-sample forecasting results.[13]

2.2 Out-of-Sample Forecasting Results

Before reporting the complete results for each of the three out-of-sample periods, following Welch and Goyal (2008), we present time-series plots of the differences between the cumulative square prediction error for the historical average benchmark forecast and the cumulative square prediction error for the forecasts based on the individual predictive regression models in Figure 1 for 1965:1–2005:4. This is an informative graphical device that provides a visual impression of the consistency of an individual predictive regression model's out-of-sample forecasting performance over time. When the curve in each panel of Figure 1 increases, the predictive regression model outperforms the historical average, while the opposite holds when the curve decreases.[14] The plots conveniently illustrate whether an individual predictive regression model has a lower MSPE than the historical average for any particular out-of-sample period by redrawing the horizontal zero line to correspond to the start of the out-of-sample period. Essentially, we compare the height of the curve at the two points corresponding to the beginning and end of a given out-of-sample period: if the curve is higher (lower) at the end of the out-of-sample period than at the beginning, the predictive regression model (historical average) has a lower MSPE over the out-of-sample period. A predictive regression model that always outperforms the historical average for any out-of-sample period will thus have a curve with a slope that is always positive; the closer a predictive regression model is to this ideal, the greater its ability to consistently beat the historical average in terms of MSPE.

The solid lines in Figure 1 illustrate that none of the 15 individual economic variables consistently outperforms the historical average. Some of the panels have positively sloped curves during certain periods, but all panels also display relatively extended periods where the curves are negatively sloped, often substantially so. It is interesting to note that a number of valuation ratios, such as the dividend-price ratio, dividend yield, and book-to-market ratio, have distinctly positive slopes in the early 1970s; the curves, however, become negatively sloped after the Oil Shock, and they become markedly negatively sloped during the 1990s. This erratic out-of-sample performance renders these valuation ratios unreliable out-of-sample predictors of the equity premium. Overall, Figure 1 visually conveys the primary message of Welch and Goyal (2008): it is difficult to identify individual predictors that reliably outperform the historical average with respect to forecasting the equity premium.

Campbell and Thompson (2008) show that imposing theoretically motivated restrictions on individual predictive regression models can improve their out-of-sample performance. We illustrate the effects of these types of restrictions with the dotted lines in Figure 1. More specifically, we set β_i to zero when recursively estimating (1) if the estimated slope does not match the theoretically expected sign; we also set the individual forecast to zero if the predictive regression model generates a negative equity premium forecast. Imposing Campbell and Thompson (2008) restrictions on the individual predictive regression model forecasts improves the out-of-sample performance of some variables for some periods, such as TBL and LTY during the 1970s, as well as D/P and D/Y toward the end of the out-of-sample period. Even after imposing restrictions, however, it remains difficult to identify individual predictors that consistently outperform the historical average.

The solid lines in Figure 2 plot the differences between the cumulative square prediction error for the historical average forecast and the cumulative square prediction error for the combination forecasts. In contrast to Figure 1, the curves in Figure 2 have slopes that are predominantly positive, indicating that the combination forecasts deliver out-of-sample gains on a considerably more consistent basis over time than the individual predictive regression models. The curves in Figure 2 are often strongly positively sloped from approximately 1965–1975, more moderately but still consistently positively sloped from the mid 1970s to the mid 1990s, slightly negatively sloped for a brief interval during the late 1990s, and positively sloped thereafter. Most notably, Figure 2 avoids the frequent, often persistent, and substantial falloffs in the curves that plague the individual models in Figure 1. This highlights that forecast combination is an effective strategy for equity premium prediction, especially compared to individual predictive regression models.

The dotted lines in Figure 2 display the results for combination forecasts based on individual predictive regression model forecasts with Campbell and Thompson (2008) restrictions imposed. The restrictions have relatively little effect in Figure 2, perhaps due to the fact that combination forecasts always satisfy the theoretical restrictions (even though the individual forecasts do not). We discuss the economic plausibility of combination forecasts in more detail in Section 4.1 below.

We turn next to the detailed results for the three out-of-sample periods, which are presented in Table 1. The table reports R^2_{OS} statistics and average utility gains for each of the individual predictive regression models and combining methods relative to the historical average benchmark model. For R^2_{OS} statistics greater than zero, statistical significance is assessed with the Clark and West (2007) *MSPE-adjusted* statistic, as discussed in Section 1.3 above.

Panel A of Table 1 reports results for the "long" 1965:1–2005:4 out-of-sample period. The second and fifth columns of Panel A reveal that only five of the 15 individual predictors have a positive R^2_{OS}, four of which are less than or equal to 0.36%; I/K is the only predictor with an R^2_{OS} greater than 0.36% (1.44%). Three of the positive R^2_{OS} statistics are significantly greater than zero at the 10% level, while only the R^2_{OS} for I/K is significant at the 5% level. The average utility gains in the third and sixth columns of Panel A generally provide greater support for out-of-sample predictability, as 13 of the 15 predictors produce positive utility gains relative to the historical average.

Turning to the results for the combination forecasts in Panel A, the most striking result is the relatively high R^2_{OS} generated by each of the combining methods. All of the R^2_{OS} statistics for the combination forecasts are greater than 3%—four of the five are greater than or equal to 3.49%—and all of the R^2_{OS} statistics are significant at the 1% level. It is interesting to observe that all of the R^2_{OS} statistics for the combining methods are greater than the largest R^2_{OS} (1.44% for I/K) among all of the individual predictors. The utility gains associated with the combination forecasts are also sizable, with four of the five combining methods yielding utility gains well above 2%. With the exception of the median, the various combining methods produce very similar forecasting results for the 1965:1–2005:4 out-of-sample period. The mean combination forecast uses the simple "$1/N$" rule that sets the combining weight to 1/15 on each individual predictive regression model forecast. While they allow for unequal and time-varying weights, it turns out that the DMSPE combination forecasts select weights relatively close to the $1/N$ rule over time.[15]

Welch and Goyal (2008) find that the out-of-sample predictive ability of many individual economic variables deteriorates markedly over the 1976–2005 out-of-sample period, and Panel B of Table 1 generally confirms this finding. Only one of the R^2_{OS} statistics (for $NTIS$) is positive for the individual predictors in the second and fifth columns of Panel B, but this value is only 0.10% (and is not significant at conventional levels). Moreover, many of the negative R^2_{OS} statistics for the individual predictors are large in terms of absolute value, so the historical average outperforms these predictors by a substantial margin. Only six of the 15 individual predictors produce a positive utility gain (see the third and sixth columns of Panel B).

Despite the poor general performance of the individual predictors in Panel B, all of the combining methods deliver positive gains over the 1976:1–2005:4 out-of-sample period. From the eighth column of Panel B, we see that the R^2_{OS} statistics for the combining methods range from 1.01%–1.51%, and all are significant at least at the 10% level. All of the utility gains are positive for the combining methods in the ninth column of Panel B, and four of the five are greater than or equal to 0.53%.

Panel C of Table 1 reports results for the 2000:1–2005:4 out-of-sample period. The second and fifth columns of Panel C indicate that among the individual predictors, most of the valuation ratios and I/K substantially outperform the historical average over this recent period, with some R^2_{OS} statistics reaching higher than 10%. The R^2_{OS} statistics for D/P, D/Y, and E/P (I/K) are significant at the 10% (5%) level. While the valuation ratios and I/K have positive and sizable R^2_{OS} statistics in Panel C, most of the remaining individual predictors have negative R^2_{OS} statistics, and many of them are large in absolute value, signaling that the predictors are substantially outperformed by the historical average. A number of the valuation ratios as well as I/K also yield sizable utility gains (see the third and sixth columns of Panel C).

Following the trend in Panels A and B, the last two columns of Panel C show that the combination forecasts typically outperform the historical average by a sizable margin. All of the R^2_{OS} statistics are positive for the combining methods in the eighth column of Panel C, and four of the five are greater than or equal to 2.56%. In addition, they are all significant at conventional levels (most are significant at the 5% level). The utility gains for the combining methods are also all positive in the ninth column of Panel C, and most are near or above 2%.[16]

Table 1 also reports results for combination forecasts based on individual predictive regression models with Campbell and Thompson (2008) restrictions imposed. For brevity, we only report results for the mean combining method, labeled as "Mean, CT" in Table 1.[17] As in Figure 2, imposing theoretically motivated restrictions on the individual predictive regression models before combining forecasts has relatively little influence.

The key findings and implications in Figures 1 and 2 and Table 1 can be summarized as follows:

- The results in Figure 1 and the first six columns of Table 1 reinforce the findings of Welch and Goyal (2008) and demonstrate that it is very difficult to identify individual economic variables capable of generating reliable out-of-sample forecasts of the equity premium. Indeed, there is no single variable among the 15 considered that delivers a positive R^2_{OS} over each of the out-of-sample periods examined in Table 1.

- Nevertheless, forecast combination outperforms the historical average by statistically and economically meaningful margins for a variety of out-of-sample periods. We have thus identified effective methods for forecasting the equity premium based on economic variables that consistently beat the historical average in real time.

3 Statistical Explanations for the Benefits of Combining

We next provide statistical explanations for the relatively good out-of-sample performance of forecast combination with respect to the equity premium. Via forecast encompassing tests, we demonstrate that combining incorporates useful forecasting information from a variety of economic variables. We also show that forecast combination reduces forecast variance and stabilizes the individual forecasts, thereby improving forecasting performance in terms of an MSPE metric. In addition, we analyze a "kitchen sink" model considered by Welch and Goyal (2008) to provide a multiple regression interpretation for the mean combination method, and we discuss alternative approaches to incorporating information from a large number of economic variables.

3.1 Forecast Encompassing Test Results

Forecast encompassing was developed by Chong and Hendry (1986) and Fair and Shiller (1990), among others, and provides a means for comparing the information content in different forecasts. Consider forming an optimal composite forecast of r_{t+1} as a convex combination of the forecasts from models i and j:

$$\hat{r}^{*}_{t+1} = (1-\lambda)\hat{r}_{i,t+1} + \lambda \hat{r}_{j,t+1}, \tag{13}$$

where $0 \leq \lambda \leq 1$. If $\lambda = 0$, then the model i forecast encompasses the model j forecast, as model j does not contain any useful information—beyond that already contained in model i—for the formation of an optimal composite forecast. In contrast, if $\lambda > 0$, then the model i forecast does not encompass the model j forecast, so model j does contain information useful for forming the optimal composite forecast (again, beyond the information already contained in model i). In essence, if we reject the null hypothesis of encompassing, then it is useful to combine forecasts from models i and j instead of relying solely on the model i forecast.

Harvey, Leybourne, and Newbold (1998, HLN) develop a statistic to test the null hypothesis that the model i forecast encompasses the model j forecast ($H_0: \lambda = 0$) against the (one-sided) alternative hypothesis that the model i forecast does not encompass the model j forecast ($H_1: \lambda > 0$). Define $d_{t+1} = (\hat{u}_{i,t+1} - \hat{u}_{j,t+1})\hat{u}_{i,t+1}$, where $\hat{u}_{i,t+1} = r_{t+1} - \hat{r}_{i,t+1}$ and $\hat{u}_{j,t+1} = r_{t+1} - \hat{r}_{j,t+1}$. Letting $\bar{d} = [1/(q-q_0)]\sum_{k=q_0+1}^{q} d_{R+k}$, the modified version of the HLN statistic can be expressed as

$$MHLN = [(q-q_0-1)/(q-q_0)][\hat{V}(\bar{d})^{-1/2}]\bar{d}, \tag{14}$$

where $\hat{V}(\bar{d}) = (q-q_0)^{-1}\hat{\phi}_0$ and $\hat{\phi}_0 = (q-q_0)^{-1}\sum_{k=q_0+1}^{q}(d_{R+k} - \bar{d})^2$. HLN recommend using the *MHLN* statistic and the t_{q-q_0-1} distribution to assess statistical significance.

Table 2 reports p-values for the *MHLN* statistic applied to the 1965:1–2005:4 out-of-sample forecasts. Each entry in the table corresponds to the null hypothesis that the forecast given in the column heading encompasses the forecast indicated in the row heading. The frequent inability of individual predictive regression model forecasts to encompass forecasts from other individual models stands out in Table 2. For example, consider the D/P ratio in the second column of Table 2. While, perhaps not surprisingly, we cannot reject the null hypothesis that the D/P forecast encompasses the forecasts for the other valuation ratios, the D/P forecast fails to encompass the forecasts for any of the other economic variables (with the exception of DFY) at the 10% significance level. Similar results hold for the other individual economic variables: each economic variable does not encompass the forecasts for at least three of the remaining variables. These encompassing tests thus indicate that it is worthwhile to combine forecasts from individual models to incorporate additional information, helping to explain the out-of-sample gains corresponding to forecast combination documented in Section 2.2 above. Finally, observe that the combination forecasts are able to encompass the forecasts from the individual predictive regression models and other combining methods. The combining methods thus incorporate the relevant information from all of the individual economic variables.

3.2 Forecast Stabilization

Analogous to including additional assets in a portfolio to reduce the portfolio's variance, combining individual forecasts helps to reduce forecast variability. Of course, combination forecast variance tends to decrease the more we diversify across individual forecasts that are weakly or negatively correlated. Table 3 shows the correlation matrix for the individual predictive regression model forecasts that form the combination forecasts. Not surprisingly, the correlations between forecasts generated by the various valuation ratios are relatively large. Many of the other correlations, however, are quite small, and a number of them are negative. These empirical facts indicate that combining forecasts is likely to reduce the variance of the combination forecasts relative to each of the individual prediction regression model forecasts. In fact, as long as the combination forecasts do not have substantial biases compared to the individual forecasts, this can reduce the MSPE, as the MSPE includes the forecast variance and squared forecast bias. The reduction in forecast variance—again, as long as it does not come at the expense of a large increase in bias—also helps the combination forecasts to outperform the historical average forecast.[18]

Figure 3 depicts individual predictive regression model forecasts for the 1965:1–2005:4 out-of-sample period. The figure also shows the mean combining method forecast in the lower-right corner. The other combination forecasts are similar to the mean forecast, and we omit them to conserve space. Figure 3 confirms that forecast combination reduces forecast variability. The individual predictive regression model forecasts are often highly variable and, as we discuss further in Section 4.1 below, imply implausibly negative or unrealistically large values for the expected equity premium. Overall, the individual forecasts appear to contain substantial "noise" and give too many false signals, hurting forecasting performance. In contrast, the mean combination forecast is more stable than the individual forecasts and exhibits more plausible fluctuations in terms of its magnitude.

Figure 4 is a scatterplot depicting the forecast variance and squared forecast bias for the individual predictive regression models, historical average, and mean combining method for the 1965:1–2005:4 out-of-sample period. Since the points corresponding to the other combining methods lie close to the mean combining method, they are not included to avoid cluttering the diagram. The scatterplot depicts how forecast combination outperforms each individual model and the historical average benchmark according to an MSPE criterion. The mean combination forecast has a lower forecast variance than all of the individual predictive regression models, confirming the visual impression from Figure 3. In addition, the mean combination forecast has a relatively small squared forecast bias, close to the smallest squared biases of the individual predictive regression models. The low forecast variance and relatively small bias enable the mean combination forecast to deliver a higher R^2_{OS} (that is, smaller MSPE) than any of the individual predictive regression models over the 1965:1–2005:4 period, as shown in Panel A of Table 1.

Figure 4 also illustrates that the mean combination forecast has a variance that is only moderately higher than that of the historical average forecast, while it has a squared bias that is substantially below the historical average. This enables the mean combination forecast to achieve a sizable reduction in MSPE relative to the historical average in Panel A of Table 1. Intuitively, combining individual predictive regression model forecasts improves forecasting performance in two ways. First, combining generates a forecast with a variance near that of the smooth historical average forecast, thereby reducing the noise in the individual predictive regression model forecasts. Second, combining incorporates information from a host of economic variables—information not contained in the historical average, which ignores economic variables—and this leads to forecasts with a substantially smaller bias than the historical average.

3.3 "Kitchen Sink" Model

Following Welch and Goyal (2008), we also consider a "kitchen sink" model that includes all 15 economic variables together in a multiple predictive regression model:

$$r_{t+1} = \alpha^{KS} + \beta_1^{KS} x_{1,t} + \cdots + \beta_N^{KS} x_{N,t} + \varepsilon_{t+1}. \tag{15}$$

Similar to Welch and Goyal (2008), however, the kitchen sink model does not perform well over the different out-of-sample periods. It has R^2_{OS} statistics of -19.35%, -35.50%, and -2.29% for the 1965:1–2005:4, 1976:1–2005:4, and 2000:1–2005:4 periods, respectively, indicating that the kitchen sink model has a higher MSPE than the historical average during each out-of-sample period. It is interesting that the kitchen sink model performs so much worse than the combining methods, since both approaches are based on the same 15 economic variables and involve estimating 15 slope coefficients. We now examine links between the mean combination and kitchen sink model forecasts to understand the superior performance of the former.[19]

For transparency, we take the average of (15) and subtract it from (15) to express the kitchen sink model in deviation form:

$$r_{t+1} - \bar{r} = \beta_1^{KS}(x_{1,t} - \bar{x}_1) + \cdots + \beta_N^{KS}(x_{N,t} - \bar{x}_N) + \varepsilon_{t+1}, \tag{16}$$

where $\bar{r}$ is the mean of r_{t+1} over the estimation period and $\bar{x}_i$ is the mean of x_i for $i = 1, \ldots, N$. In matrix notation,

$$\tilde{r} = \tilde{X}\beta^{KS} + \varepsilon, \tag{17}$$

where $\tilde{r}$ is an m-vector of demeaned r_{t+1} observations available over the estimation period, $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_N)$, $\tilde{x}_i$ is an m-vector of demeaned $x_{i,t}$ observations for $i = 1, \ldots, N$, $\beta^{KS} = (\beta_1^{KS}, \ldots, \beta_N^{KS})'$, and ε is an m-vector of disturbance terms. Without loss of generality, we standardize the $\tilde{x}_i$ variables: $\tilde{x}_i'\tilde{x}_i = 1$ for $i = 1, \ldots, N$. Numerically, estimating β_i^{KS} for $i = 1, \ldots, N$ from either (15) or (16) produces the same result. A forecast of r_{t+1} based on (16) can be computed by adding back $\bar{r}$. The unrestricted OLS estimator of the kitchen sink model is given by

$$\hat{\beta}^{KS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{r}, \tag{18}$$

and the kitchen sink model forecast of r_{t+1} can be expressed as

$$\hat{r}_{t+1}^{KS} = \bar{r} + \sum_{i=1}^{N} \hat{\beta}_i^{KS}(x_{i,t} - \bar{x}_i). \tag{19}$$

One potential reason for the poor performance of $\hat{r}_{t+1}^{KS}$ is that the $N \times N$ covariance matrix, $\tilde{X}'\tilde{X}$, does not have a well-defined inverse due to collinearity. This motivates imposing restrictions on $\tilde{X}'\tilde{X}$.[20]

Suppose we restrict $\tilde{X}'\tilde{X}$ to be diagonal or $\tilde{x}_i'\tilde{x}_j = 0$ for $i \neq j$. The restricted estimator is given by

$$\hat{\beta}^D = [\mathrm{diag}(\tilde{X}'\tilde{X})]^{-1}\tilde{X}'\tilde{r} = \hat{\beta}, \tag{20}$$

where $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_N)'$, the vector of slope coefficients from the N individual bivariate regression models. The forecast corresponding to this restricted multiple regression model can be expressed as

$$\hat{r}_{t+1}^{D} = \bar{r} + \sum_{i=1}^{N} \hat{\beta}_i(x_{i,t} - \bar{x}_i) = -(N-1)\bar{r} + \sum_{i=1}^{N}(\hat{\alpha}_i + \hat{\beta}_i x_{i,t}), \tag{21}$$

where we use the well-known bivariate regression result that $\hat{\alpha}_i = \bar{r} - \hat{\beta}_i\bar{x}_i$. Comparing (21) to (19), assuming that $\tilde{X}'\tilde{X}$ is diagonal entails replacing the multiple regression slope coefficient estimates with their bivariate counterparts. While this forecast is based on the individual bivariate slope coefficient estimates, it still differs from the mean combination forecast.

Alternatively, suppose we impose an even stronger restriction on each of the multiple regression slope coefficients in (16):

$$\beta_i^{KS} = \frac{1}{N}\beta_i, \qquad i = 1, \ldots, N. \tag{22}$$

The corresponding forecast is given by

$$\hat{r}_{t+1} = \bar{r} + \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_i(x_{i,t} - \bar{x}_i) = \frac{1}{N}\sum_{i=1}^{N}(\bar{r} - \hat{\beta}_i\bar{x}_i) + \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i x_{i,t} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_i + \hat{\beta}_i x_{i,t}), \tag{23}$$

which is clearly the mean combination forecast. This shows that the mean combination forecast can be viewed as a restricted forecast from a multiple regression or kitchen sink model. To provide some basic intuition for the restriction in (22), consider the simple case of $N = 2$, and assume $x_{1,t}$ and $x_{2,t}$ are correlated to some extent—as predictors are likely to be in actual data. Suppose the two bivariate forecasting models generate r_{t+1} forecasts of $\bar{r} + \hat{\beta}_1(x_{1,t} - \bar{x}_1) = \bar{r} + 0.02$ and $\bar{r} + \hat{\beta}_2(x_{2,t} - \bar{x}_2) = \bar{r} + 0.04$, respectively. In the context of a multiple regression that includes both $\hat{\beta}_1(x_{1,t} - \bar{x}_1)$ and $\hat{\beta}_2(x_{2,t} - \bar{x}_2)$, we do not want to simply add 0.02 and 0.04 and use $\bar{r} + 0.06$ as the forecast, since this is likely to produce a substantially biased forecast. Instead, treating each variable equally, we can reduce the bias by scaling down the slope coefficient of each variable by a factor of $1/2$, resulting in a forecast of $\bar{r} + 0.03$. In the special case where both bivariate forecasts are unbiased, the scaled forecast must also be unbiased.

The restriction that $\beta_i^{KS} = (1/N)\beta_i$ is obviously very strong. Why does the combination method work well in our (and macroeconomic) applications? As we argued in the introduction, the data-generating process for the equity premium is highly complex and constantly evolving, with individual variables providing accurate signals during some periods but numerous false signals during others. This confounds estimating the unrestricted kitchen sink model and seriously compromises its forecasting ability.[21] Examining (23), we see that the mean combining method "shrinks" the forecast toward the historical average, and this is evident in Figure 5, which depicts the kitchen sink, mean combination, and historical average forecasts. The kitchen sink model forecast is much more volatile than the mean combination forecast, and, looking back to Figure 3, more volatile than the individual predictive regression model forecasts (note the difference in the scales of the vertical axes in Figures 3 and 5).[22] In the same way discussed in Section 3.2 above for the individual predictive regression model forecasts, the $\beta_i^{KS} = (1/N)\beta_i$ restriction implicit in the mean combination forecast stabilizes the kitchen sink model forecast, while still incorporating meaningful information from all of the economic variables. The stabilization afforded by the combining approach substantially improves out-of-sample equity premium forecasts.

Another approach for incorporating information from economic variables is factor analysis. This approach involves extracting a relatively small number of common factors from a larger number of variables and using these factors in a single forecasting model; Huang and Lee (2009) call this "combination of information."[23] Huang and Lee (2009) provide an interesting comparison of forecast combination and information combination, and they find that forecast combination typically outperforms information combination with respect to forecasting the equity premium using a common set of twelve potential predictors. Ludvigson and Ng (2007) employ a common factor approach based on an extremely large number (350) of macroeconomic and financial variables, and they find considerable in-sample predictive ability for the factors. While Ludvigson and Ng (2007) detect statistically significant out-of-sample predictive power, they do not focus on real-time out-of-sample forecasting.[24] We concentrate on individual predictive regression models based on 15 well-known economic variables from the predictability literature and combinations of forecasts generated by these models, instead of the 350 variables considered by Ludvigson and Ng (2007). Our paper's combination and Ludvigson and Ng (2007) factor approaches represent different strategies for utilizing a range of information sets. It will be interesting in future research to compare these approaches, building on the analysis in Huang and Lee (2009), and to explore potential gains to using them in conjunction.

4 Links to the Real Economy

In this section, we investigate links between equity premium forecasts and the real economy. Such links provide additional support for the combination approach to equity premium prediction, and they provide an economic rationale for the out-of-sample gains associated with combination forecasts.

4.1 Equity Premium Forecasts and NBER-Dated Business-Cycle Phases

Fama and French (1989) and Cochrane (1999, 2007) argue that heightened risk aversion during economic downturns demands a higher risk premium, thereby generating equity premium predictability.[25] In light of this, we examine fluctuations in combination forecasts of the equity premium over the business cycle. More specifically, we show that movements in combination forecasts are closely connected to NBER-dated business-cycle phases.

Figure 6 depicts the mean combination forecast of the equity premium, along with vertical lines indicating NBER-dated business-cycle peaks and troughs. There are six recessions over the 1965:1–2005:4 out-of-sample period, with business-cycle peaks (troughs) occurring at 1969:4 (1970:4), 1973:4 (1975:1), 1980:1 (1980:3), 1981:3 (1982:4), 1990:3 (1991:1), and 2001:1 (2001:4). Figure 6 shows well-defined patterns in the mean combination forecast around these peaks and troughs. There are distinct upward spikes in the combination forecast at or shortly after the troughs associated with the four relatively deep recessions of the 1970s and early 1980s; indeed, the combination forecast takes four of its highest values over the out-of-sample period near these troughs.[26] Furthermore, the combination forecast takes four of its lowest values very near the peaks preceding these recessions. In general, we see declines in the equity premium forecast during expansions and sharp increases during recessions. Observe that the increases in the equity premium forecast during the 1990–1991 and 2001 recessions are much more modest than those during the earlier four recessions. This makes sense, since the latter two recessions were much milder relative to the earlier four recessions. Overall, Figure 6 demonstrates that the combination approach produces an equity premium forecast that closely tracks NBER business-cycle phases, and the behavior of the forecast agrees with the Fama and French (1989) and Cochrane (1999, 2007) account of equity premium predictability.

Looking back to Figure 3, it is instructive to compare fluctuations in the individual predictive regression model and mean combination forecasts from an economic perspective. As we discussed in Section 3.2 above, the combination forecast substantially reduces the volatility of the individual forecasts. This is important from an economic perspective, since the individual models frequently generate implausibly large fluctuations in the equity risk premium. For example, a number of the individual predictive regression models produce equity premium forecasts between approximately 0.06 and 0.10, implying an annual equity risk premium ranging from 24% to 40%, which seems implausibly large. Furthermore, a number of the individual models predict a negative premium—sometimes falling to -20% on an annual basis—during certain periods. As argued by Campbell and Thompson (2008), a negative equity premium is economically implausible. In contrast to the mean combination forecast, the individual forecasts also appear less closely related to the NBER business-cycle phases. The combination approach thus produces a highly plausible out-of-sample measure of a time-varying equity risk premium by stabilizing individual predictive regression model forecasts and better connecting them to the business cycle.[27]

While the individual predictive regression model forecasts often exhibit economically implausible fluctuations, the historical average forecast, which is based on the constant expected equity premium model, appears too smooth. That is, from an economic perspective, the "problem" with the historical average forecast is that it ignores business-cycle fluctuations and thus fails to incorporate meaningful macroeconomic information. Overall, Figures 3 and 6 indicate that the combination approach includes relevant macroeconomic information missed by the historical average forecast while avoiding the implausible fluctuations in the equity risk premium associated with individual predictive regression models.

As indicated in Table 4, combination forecasts of the equity premium for time t are also significantly correlated with growth rates in three macroeconomic variables—real GDP, real profits, and real net cash flows—at time t.[28] These are three highly relevant aggregates for the equity market. Table 4 shows that combination forecasts are all positively and significantly correlated with growth rates in the three macroeconomic variables, and many of the correlations are near or greater than 0.30. These correlations complement the evidence in Figure 6 and further demonstrate that the combination forecasts are related to the real economy.[29]

4.2 Forecasting Gains During "Good" and "Bad" Growth Periods

In the spirit of Liew and Vassalou (2000), we next analyze combination forecasts during periods of "good," "normal," and "bad" economic growth.[30] More specifically, we compute R^2_{OS} statistics during good, normal, and bad growth regimes, where the regimes are based on sorted values of real GDP, real profit, and real net cash flow growth (in turn). To ensure that we have a reasonable number of observations in each regime, good, normal, and bad periods are defined using the top, middle, and bottom third of sorted growth rates, respectively. The results are reported in Table 5. To examine the robustness of the results, we also analyze equity premium forecasts at the four-quarter horizon. The combination forecasts at the four-quarter horizon are based on individual predictive regression models that are a straightforward extension of (1):

$$r_{t+1:t+4} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1:t+4}, \tag{24}$$

where $r_{t+1:t+4} = r_{t+1} + \cdots + r_{t+4}$ and the forecasts are again computed recursively as described in Section 1.1 above. The historical average forecast simply sets $\beta_i = 0$ in (24). Due to overlapping observations, it is necessary to allow for autocorrelation when computing the Clark and West (2007) *MSPE-adjusted* statistic to assess the significance of the R^2_{OS} statistic, and we use the Newey and West (1987) standard error estimate.

Intuitively, we expect out-of-sample gains for an equity premium forecast linked to macroeconomic fundamentals to be particularly manifest during more extreme periods characterized by relatively high or low growth. Table 5 shows that out-of-sample gains for the combination forecasts are often concentrated in extreme periods, especially periods of low growth. At the one-quarter horizon, the R^2_{OS} statistics are always higher during low-growth compared to normal-growth periods; sorting on real GDP growth, the R^2_{OS} statistics are approximately four times higher during low-growth compared to normal-growth periods. These differences are magnified when we sort on real profit growth, while they are muted somewhat when we sort on real net cash flow growth. The R^2_{OS} statistics for the one-quarter horizon are also all higher during high-growth relative to normal-growth periods. The higher R^2_{OS} statistics during high-growth periods are especially evident when we sort on real net cash flow growth.

Comparing the second and sixth columns of Table 5, the R^2_{OS} statistics typically more than double over the 1965:1–2005:4 out-of-sample period at the four-quarter horizon relative to the one-quarter horizon. Matching the one-quarter-horizon results, comparing the eighth and ninth columns of Table 5 shows that the R^2_{OS} statistics are all greater during low-growth compared to normal-growth periods, and the differences are again particularly evident when we sort on real GDP and real profit growth. At the four-quarter horizon, there is less evidence of increased out-of-sample gains during high-growth compared to normal-growth periods, although it is still evident for most combining methods when we sort on real GDP growth. Overall, Table 5 points to enhanced out-of-sample gains for the combining methods relative to the historical average during extreme—especially low-growth—periods, again linking combination forecasts to important macroeconomic fluctuations.

4.3 Forecasting Macroeconomic Growth with the Same Set of 15 Economic Variables

Cochrane (2007) stresses that equity premium forecasts are more plausibly related to macroeconomic risk if equity premium predictors can also forecast business cycles, and, indeed, there is evidence that some equity premium predictors from the literature have predictive ability with respect to real output growth; see, for example, Estrella and Hardouvelis (1991), Harvey (1989, 1993), and Ang, Piazessi, and Wei (2006) with respect to the term spread.

As shown by Stock and Watson (2003) and others, however, the forecasting power of individual economic variables with respect to output growth can be highly unstable over time, very similar to the situation for out-of-sample equity premium prediction documented by Welch and Goyal (2008) and in Section 2.2 above. Interestingly, Stock and Watson (2003) also show that combination forecasts of output growth consistently outperform an autoregressive (AR) benchmark model. This provides a potential explanation for the out-of-gains associated with combination forecasts of the equity premium: individual economic variables fail to consistently generate out-of-sample gains with respect to equity premium prediction because they produce erratic gains with respect to forecasting macroeconomic fluctuations; in contrast, forecast combination produces consistent out-of-sample gains for equity premium prediction because it also produces steady gains for predicting macroeconomic fluctuations. We provide support for this explanation for the set of 15 economic variables considered in the present paper.

Similar to Stock and Watson (2003), we form macroeconomic growth forecasts using the following autoregressive distributed lag (ARDL) model:

$$y_{t+1} = \zeta_i + \eta_i y_t + \lambda_i x_{i,t} + \upsilon_{t+1}, \tag{25}$$

where y_{t+1} is the growth rate of real GDP, real profits, or real net cash flows from period t to $t+1$. We simply replace y_{t+1} with $y_{t+1:t+4} = y_{t+1} + \cdots + y_{t+4}$ in (25) to consider a four-quarter horizon. The lagged y_t term is included in (25) to accommodate the autocorrelation in y_t.[31] We generate out-of-sample forecasts of y_{t+1} by estimating (25) recursively, analogous to the procedure described in Section 1.1 above for individual equity premium predictive regression models. Following the recent macroeconomic forecasting literature, an AR model, (25) with $\lambda_i = 0$, serves as a natural benchmark forecasting model. We compute a suitably modified version of the R^2_{OS} statistic to measure the reduction in MSPE for the ARDL model relative to the AR benchmark, as well as the Clark and West (2007) statistic to assess the statistical significance of R^2_{OS}. We also compute combination forecasts for y_{t+1} and $y_{t+1:t+4}$ using the same set of combining methods described in Section 1.2 above.

The results are reported in Table 6 for the 1965:1–2005:4 and 1976:1–2005:4 out-of-sample periods. Panel A shows that the individual economic variables typically fail to outperform the AR benchmark model, sometimes by a substantial margin. None of the individual economic variables produces a positive R^2_{OS} for all three macroeconomic variables over both out-of-sample periods. Panel B of Table 6 shows that the combination forecasts almost always generate significant and substantial out-of-sample gains for all three macroeconomic variables.[32] The results in Table 6 are parallel to those in Table 1. As suggested above, the parallels between forecasting the equity premium and business-cycle fluctuations help to explain the success of forecast combination by directly linking it to the real economy: forecast combination based on 15 individual economic variables from the literature improves out-of-sample equity premium prediction because forecast combination also improves prediction of macroeconomic fluctuations based on the same 15 economic variables.

4.4 Structural Breaks in Macroeconomic Relationships and Related Breaks in Equity Premium Predictive Regression Models

A natural explanation for the inconsistent out-of-sample performance of individual predictive regression models, and one stressed by Welch and Goyal (2008), is structural instability. Figure 7 gives a visual impression of the changing nature of the relationships between the equity premium and the individual economic variables over the 1947:3–2005:4 period. The figure depicts correlations between r_{t+1} and $x_{i,t}$ $(i = 1, \ldots, 15)$ calculated on the basis of 10-year moving windows of data. The correlations fluctuate substantially over the postwar period, and there are numerous instances where the correlation moves from being significant during certain periods to insignificant during others. Overall, Figure 7 suggests important structural instabilities in the relationships between the equity premium and 15 economic variables, complementing recent empirical evidence of structural breaks in individual equity premium predictive regression models in Paye and Timmermann (2006) and Rapach and Wohar (2006a).[33] As discussed above, structural instabilities in individual predictive regression models help to explain the success of forecast combination for out-of-sample equity premium prediction, since forecast combination can improve the performance of individual forecasting models in the presence of structural breaks [Hendry and Clements (2004), Clements and Hendry (2006), Timmermann (2006)].

Structural instabilities in macroeconomic relationships potentially underlie the structural instabilities in individual equity premium predictive regression models. We examine ties between structural breaks in macroeconomic relationships and equity premium predictive regression models in the following manner. First, we use the Bai and Perron (1998) methodology to test for (potentially multiple) structural breaks in individual predictive regression models of real GDP growth. Second, a Chow test determines whether the structural breaks indicated by the Bai and Perron (1998) methodology in the first step are also significant in individual predictive regression models of the equity premium.

The predictive regression model for real GDP growth is given by

$$y_{t+1} = \zeta_i + \lambda_i x_{i,t} + \upsilon_{t+1}, \tag{26}$$

which is similar to (25), except that it excludes the lagged y_t term. We omit this term in (26) to give it the same basic structure as (1); we instead account for autocorrelation in real output growth by allowing for autocorrelation in the disturbance term, υ_{t+1}. We employ the Bai and Perron (1998) $UDmax$ and $WDmax(10\%)$ statistics to test for the existence of one or more breaks in (26).[34] More specifically, the statistics are used to test the null hypothesis of no breaks against the alternative hypothesis of one to eight breaks. Details of the computation of the statistics are provided in Bai and Perron (1998, 2003). Note that the tests require a minimum length for each regime, and following the recommendation of Bai and Perron (2003), we assume a minimum length equal to 10% of the sample size given that we allow for a maximum of eight breaks. Bai and Perron (2003) advise using the *Dmax* statistics to first determine whether any breaks exist, and then examining a sequence of $F(l+1|l)$ statistics to determine the number of breaks. Details on the computation of the $F(l+1|l)$ statistics and how they can be used to select the number of breaks are given in Bai and Perron (2003). They show that this strategy has reasonable size and power properties. After determining the number of breaks, the Bai and Perron (1998) algorithm estimates the location of the breaks and model parameters for each regime.

Table 7 reports $UDmax$ and $WDmax(10\%)$ statistics and estimated break dates for individual predictive regression models of real GDP growth based on the 15 economic variables considered in the present paper.[35] There is extensive evidence of structural instability in the individual predictive regression models of real GDP growth, and both the $UDmax$ and $WDmax(10\%)$ statistics in the second and third columns are significant at conventional levels for 14 of the 15 individual models. A number of the breaks occur near the mid 1970s, corresponding to the Oil Shocks, and mid 1980s, shortly after the change in Federal Reserve operating procedures. In addition, around half of the models have breaks occurring during the early to mid 1950s, close to Treasury-Federal Reserve Accord and accompanying the transition from the war-time economy. The extensive evidence of structural breaks in macroeconomic relationships in Table 7 complements the evidence in Stock and Watson (1996, 2003).

The tenth column of Table 7 reports χ^2-statistics corresponding to Chow tests applied to individual equity premium predictive regression models. The break dates considered for the Chow test are those identified for the individual real GDP growth predictive regression models given in the fourth through ninth columns of Table 7. Observe that the χ^2-statistic is significant for equity premium predictive regression models based on 10 of the 14 economic variables that deliver significant $UDmax$ and $WDmax(10\%)$ statistics for the real GDP growth predictive regression models.[36] Overall, Table 7 shows that significant structural breaks in macroeconomic relationships frequently correspond to significant simultaneous breaks in equity premium predictive regression models. As we have emphasized, forecast combination helps to improve out-of-sample equity premium prediction in the presence of structural instability in individual predictive regression models. Given that structural breaks in equity premium predictive regression models are often related to breaks in macroeconomic relationships, we have further evidence of links between combination forecasts of the equity premium and the real economy.[37]

5 Conclusion

While numerous economic variables have been identified in the literature with in-sample predictive ability for the equity premium, Welch and Goyal (2008) show that individual variables fail to deliver consistent out-of-sample forecasting gains relative to the historical average. In contrast, forecast combination methods provide convincing evidence of the out-of-sample predictive ability of 15 economic variables taken as a whole over a number of periods. We thus find that the data significantly support out-of-sample return predictability using economic variables. In addition to being of practical interest, this has important theoretical implications, as recently emphasized by

Cochrane (2008), who provides a strong theoretical rationale for the predictability of stock returns.

Forecast combination appears successful for out-of-sample equity premium prediction because it achieves a middle ground. On the one hand, individual predictive regression model forecasts often appear too volatile to represent plausible changes in the expected equity premium; forecast combination can substantially reduce forecast variance. On the other hand, the historical average forecast appears too smooth, thereby ignoring information contained in economic variables that potentially affects the expected equity premium; forecast combination includes information from numerous economic variables. Overall, forecast combination thus strikes a balance by incorporating information from a host of economic variables in a way that avoids excessively noisy forecasts.

We also show that combination forecasts of the equity premium are linked to the real economy. This is important, as it puts the combination approach on a firm macroeconomic footing. We link combination forecasts of the equity premium to the real economy in the following ways. First, combination forecasts of the equity premium are closely related to NBER-dated business-cycle phases, in agreement with the Fama and French (1989) and Cochrane (1999, 2007) view that heightened risk aversion during economic downturns requires a higher risk premium. Second, out-of-sample gains accruing to combination forecasts are often concentrated in relatively extreme periods of economic growth, especially periods of low growth. This makes sense, since we expect out-of-sample gains for a time-varying equity premium forecast linked to macroeconomic fundamentals to be more manifest during more extreme growth periods. Third, the same set of 15 economic variables that generate combination forecasts of the equity premium also produce out-of-sample gains when forming combination forecasts of macroeconomic variables such as real GDP growth, real earnings growth, and real net cash flow growth. This supports the contention of Cochrane (2007), who stresses that equity premium forecasts are more plausibly related to macroeconomic risk if equity premium predictors can also forecast business cycles. Lastly, we find that structural breaks in macroeconomic relationships are frequently linked to significant breaks in equity premium predictive regression models.

Our evidence suggests that the usefulness of forecast combination methods ultimately stems from the highly uncertain, complex, and constantly evolving data-generating process underlying expected equity returns, which are related to a similar process in the real economy. This type of process will be difficult to approximate with a single, relatively parsimonious predictive regression model, while forecast combination reduces the uncertainty/instability risk associated with reliance on a single model. Our results imply that existing asset pricing models, which typically rely on one or a few state variables to determine time-varying expected returns and the associated investment opportunities, will have difficulty accurately tracking the expected equity premium over time. Future applied asset pricing models could benefit from the consideration of more complex data-generating processes with more variables that better mimic time-varying fluctuations in expected returns related to the real economy, and the combination strategy used in the present paper provides a tractable way of doing this.

Out-of-sample equity premium predictability will always be a challenging task, and we certainly do not claim that combination forecasts will outperform the historical average over every possible out-of-sample period. Nevertheless, we find that combination forecasts outperform the historical average by statistically and economically meaningful margins on a reasonably consistent basis over time, certainly much more consistently than numerous individual predictive regression models from the literature. This fact, together with the links we document between combination forecasts of the equity premium and the real economy, suggest that the forecast combination approach will remain a useful strategy for out-of-sample equity premium prediction in the future.

References

Aiolfi, M., and C.A. Favero. 2005. Model Uncertainty, Thick Modelling and the Predictability of Stock Returns. *Journal of Forecasting* 24:233–54.

Ang, A., and G. Bekaert. 2007. Return Predictability: Is it There? *Review of Financial Studies* 20:651–707.

Ang, A., M. Piazessi, and M. Wei. 2006. What Does the Yield Curve Tell Us About GDP Growth? *Journal of Econometrics* 131:359–403.

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

Bai, J., and P. Perron. 1998. Estimating and Testing Linear Models With Multiple Structural Changes. *Econometrica* 66:47–78.

Bai, J., and P. Perron. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics* 18:1–22.

Baker, M., and J. Wurgler. 2000. The Equity Share in New Issues and Aggregate Stock Returns. *Journal of Finance* 55:2219–57.

Bates, J.M., and C.W.J. Granger. 1969. The Combination of Forecasts. *Operational Research Quarterly* 20:451–68.

Bossaerts, P., and P. Hillion. 1999. Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies* 12:405–28.

Boudoukh, J., R. Michaely, M.P. Richardson, and M.R. Roberts. 2007. On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing. *Journal of Finance* 62:877–915.

Boudoukh, J., M.P. Richardson, and R.F. Whitelaw. 2008. The Myth of Long-Horizon Predictability. *Review of Financial Studies* 21:1577–605.

Breen, W., L.R. Glosten, and R. Jagannathan. 1989. Economic Significance of Predictable Variations in Stock Index Returns. *Journal of Finance* 64:1177–89.

Campbell, J.Y. 1987. Stock Returns and the Term Structure. *Journal of Financial Economics*

18:373–399.

Campbell, J.Y. 2000. Asset Pricing at the Millennium. *Journal of Finance* 55:1515–67.

Campbell, J.Y. 2008. Viewpoint: Estimating the Equity Premium. *Canadian Journal of Economics* 41:1–21.

Campbell, J.Y., and J.H. Cochrane. 1999. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107:205–51.

Campbell, J.Y., and R.J. Shiller. 1988. Stock Prices, Earnings, and Expected Dividends. *Journal of Finance* 43:661–76.

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

Campbell, J.Y., and S.B. Thompson. 2008. Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21:1509–31.

Campbell, J.Y., and T. Vuolteenaho. 2004. Inflation Illusion and Stock Prices. *American Economic Review* 94:19–23.

Chong, Y.Y., and D.F. Hendry. 1986. Econometric Evaluation of Linear Macro-Econometric Models. *Review of Economic Studies* 53:671–90.

Clark, T.E., and M.W. McCracken. 2001. Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105:85–110.

Clark, T.E., and K.D. West. 2007. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:291–311.

Clements, M.P., and D.F. Hendry. 2006. Forecasting with Breaks, in G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Amsterdam: Elsevier.

Cochrane, J.H. 1991. Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations. *Journal of Finance* 46:209–37.

Cochrane, J.H. 1999. New Facts in Finance. Federal Reserve Bank of Chicago *Economic Perspectives* 23:36–58.

Cochrane, J.H. 2007. Financial Markets and the Real Economy, in R. Mehra (ed.), *Handbook of* *the Equity Premium*. Amsterdam: Elsevier.

Cochrane, J.H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies* 21:1533–75.

Cremers, K.J.M. 2002. Stock Return Predictability: A Bayesian Model Selection Perspective. *Review of Financial Studies* 15:1223–49.

Dangl, T., and M. Halling. 2007. Predictive Regressions with Time-Varying Coefficients. Manuscript, University of Utah.

Diebold, F.X., and R.S. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13:253–263.

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

Draper, D. 1995. Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society Series B* 58:45–97.

Estrella, A., and G.A. Hardouvelis. 1991. The Term Structure as Predictor of Real Economic Activity. *Journal of Finance* 46:555–76.

Fair, R.C., and R.J. Shiller. 1990. Comparing Information in Forecasts from Econometric Models. *American Economic Review* 80:375–89.

Fama, E.F., and K.R. French. 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22:3–25.

Fama, E.F., and K.R. French. 1989. Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* 25:23–49.

Fama, E.F., and K.R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:3–56.

Fama, E.F., and G.W. Schwert. 1977. Asset Returns and Inflation. *Journal of Financial Economics* 5:115–46.

Goyal, A., and I. Welch. 2003. Predicting the Equity Premium with Dividend Ratios. *Management Science* 49:639–54.

Guo, H. 2006. On the Out-of-Sample Predictability of Stock Market Returns. *Journal of Business*

79:645–70.

Harvey, C.R. 1989. Forecasts of Economic Growth from Stock and Bond Markets. *Financial Analysts Journal* 45:38–45.

Harvey, C.R. 1993. The Term Structure Forecasts Economic Growth. *Financial Analysts Journal* 49:6–8.

Harvey, D.I., S.J. Leybourne, and P. Newbold. 1998. Tests for Forecast Encompassing. *Journal of Business and Economic Statistics* 16:254–59.

Hendry, D.F., and M.P. Clements. 2004. Pooling of Forecasts. *Econometrics Journal* 7:1–31.

Huang, H., and T.-H. Lee. 2009. To Combine Forecasts or to Combine Information. *Econometric Reviews*, forthcoming.

Kim, C.-J., J.C. Morley, and C.R. Nelson. 2005. The Structural Break in the Equity Premium. *Journal of Business and Economic Statistics* 23:181–91.

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

Lettau, M., and S. Ludvigson. 2001. Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 56:815–49.

Lettau, M., and S. Van Nieuwerburgh. 2008. Reconciling the Return Predictability Evidence. *Review of Financial Studies* 21:1607–52.

Liew, J., and M. Vassalou. 2000. Can Book-to-Market, Size, and Momentum be Risk Factors that Explain Economic Growth. *Journal of Financial Economics* 57:221–45.

Ludvigson, S.C., and S. Ng. 2007. The Empirical Risk-Return Relation: A Factor Analysis Approach. *Journal of Financial Econometrics* 83:171–222.

Mamaysky, H., M. Spiegel, and H. Zhang. 2007. Improved Forecasting of Mutual Fund Alpha and Betas. *Review of Finance* 11:359–400.

Mamaysky, H., M. Spiegel, and H. Zhang. 2008. Estimating the Dynamics of Mutual Fund Alphas and Beta. *Review of Financial Studies* 21:233–64.

Marquering, W., and M. Verbeek. 2004. The Economic Value of Predicting Stock Index Returns and Volatility. *Journal of Financial and Quantitative Analysis* 39:407–429.

McCracken, M.W. 2007. Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* 140:719–52.

Nelson, C.R. 1976. Inflation and the Rates of Return on Common Stock. *Journal of Finance* 31:471–83.

Newey, W.K., and K.D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.

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

Pástor, L., and R.F. Stambaugh. 2001. The Equity Premium and Structural Breaks. *Journal of Finance* 56:1207–39.

Pástor, L., and R.F. Stambaugh. 2008. Predictive Systems: Living with Imperfect Predictors. *Journal of Finance*, forthcoming.

Pesaran, M.H., D. Pettenuzzo, and A. Timmermann. 2006. Forecasting Time Series Subject to Multiple Structural Breaks. *Review of Economic Studies* 73:1057–84.

Pesaran, M.H., and A. Timmermann. 1995. Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance* 50:1201–28.

Pesaran, M.H., and A. Timmermann. 2005. Real-Time Econometrics. *Econometric Theory* 21:212–31.

Pettenuzzo, D., A. Timmermann, R. Valkanov, and R. Wu. 2008. Return Predictability Under Equilibrium Constraints on the Equity Premium. Manuscript, University of California at San Diego.

Pontiff, J., and L.D. Schall. 1998. Book-to-Market Ratios as Predictors of Market Returns. *Journal of Financial Economics* 49:141–60.

Rapach, D.E.. and M.E. Wohar. 2006a. Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns. *Journal of Financial Econometrics* 4:238–74.

Rapach, D.E., and M.E. Wohar. 2006b. In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining. *Journal of Empirical Finance* 13:231–47.

Spiegel, M. 2008. Forecasting the Equity Premium: Where We Stand Today. *Review of Financial Studies* 21:1453–54.

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

Stock, J.H., and M.W. Watson. 1996. Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics* 14:11–29.

Stock, J.H., and M.W. Watson. 1999. Forecasting Inflation. *Journal of Monetary Economics* 44:293–335.

Stock, J.H., and M.W. Watson. 2003. Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature* 41:788–829.

Stock, J.H., and M.W. Watson. 2004. Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* 23:405–30.

Stock, J.H., and M.W. Watson. 2006. Forecasting with Many Predictors, in G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Amsterdam: Elsevier.

Theil, H. 1971. *Applied Economic Forecasting*. Amsterdam: North Holland.

Timmermann, A. 2006. Forecast Combinations, in G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Amsterdam: Elsevier.

Timmermann, A. 2008. Elusive Return Predictability. *International Journal of Forecasting* 24:1–18.

Timmermann, A., and C.W.J. Granger. 2004. Efficient Market Hypothesis and Forecasting. *International Journal of Forecasting* 20:15–27.

Viceira, L.M. 1997. Testing for Structural Change in the Predictability of Asset Returns. Manuscript, Harvard Business School.

Wachter, J.A., and M. Warusawitharana. 2009. Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market? *Journal of Econometrics* 148:162–78.

Welch, I., and A. Goyal. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21:1455–508.

West, K.D. 1996. Asymptotic Inference About Predictive Ability. *Econometrica* 64:1067–84.

Notes

[1]The list of studies above is not meant to be exhaustive; see Campbell (2000) and Welch and Goyal (2008) for more extensive surveys of the vast literature on return predictability.

[2]See Welch and Goyal (2008), Campbell and Thompson (2008), Cochrane (2008), Boudoukh, Richardson, and Whitelaw (2008), and Lettau and Van Nieuwerburgh (2008).

[3]Campbell and Thompson (2008) find that placing theoretically motivated restrictions on individual predictive regression models helps to improve their out-of-sample performance in statistical and economic terms; also see Campbell (2008).

[4]Timmermann (2008) discusses how the changing nature of the data-generating process makes return predictability "elusive."

[5]In an environment relevant to equity premium forecasting, Hendry and Clements (2004) show that combining forecasts across individual models can lead to improved forecast accuracy when individual models provide only partial (perhaps overlapping) descriptions of the data-generating process and are subject to periodic structural breaks; also see Clements and Hendry (2006) and Timmermann (2006).

[6]We obtain qualitatively similar results using monthly data. The results for monthly data are available upon request.

[7]See Timmermann (2006) for an extensive survey of forecast combination. Also see Timmermann and Granger (2004) and Pesaran and Timmermann (2005) for interesting applications of model combination more generally.

[8]In a variation of Pesaran and Timmermann (1995), Aiolfi and Favero (2005) use a type of combining approach, "thick" modeling, in forecasting the equity premium. Timmermann (2008) considers adaptive methods for combining primarily linear and nonlinear autoregressive model forecasts of monthly stock returns.

[9]Complete results for this combining method are available upon request. Bayesian analysis of predictive regression models of stock returns in various contexts is provided by Stambaugh (1999), Avramov (2002), Cremers (2002), Dangl and Halling (2007), and Pástor and Stambaugh (2008). While beyond the scope of the present paper, which adopts a classical approach, it would be interesting in future research to incorporate these types of Bayesian techniques in out-of-sample equity premium forecasting; see Pettenuzzo, Timmermann, Valkanov, and Wu (2008) for promising research in this direction.

[10]The Diebold and Mariano (1995) and West (1996) statistic can be severely undersized when comparing forecasts from nested linear models, leading to tests with very low power. Rapach and Wohar (2006b) find that there is stronger evidence of out-of-sample predictive ability for individual economic variables with respect to stock returns when tests with good size and power are used.

[11]Following Campbell and Thompson (2008), we constrain the portfolio weight on stocks to lie between 0% and 150% (inclusive) each month, so that $w_{0,t} = 0$ ($w_{0,t} = 1.5$) if $w_{0,t} < 0$ ($w_{0,t} > 1.5$) in (9).

[12]The data are available at www.bus.emory.edu/AGoyal/Research.html.

[13]Note that the out-of-sample periods refer to the periods used to evaluate the out-of-sample forecasts. As indicated in Section 1.2 above, some of the combining methods require a holdout out-of-sample period, and we use the ten years (40 quarters) before the start of the out-of-sample evaluation period as the initial holdout out-of-sample period.

[14]As pointed out by Welch and Goyal (2008), the units on the plots are not intuitive.

[15]To conserve space, we do not report the complete set of DMSPE combining weights. They are available upon request. The DMSPE combining weights are apparently close to $1/N$ due to their reliance on forecasting performance over a holdout out-of-sample period. Because of the uncertainty/instability associated with individual models, no individual model tends to dominate for a reasonably long holdout out-of-sample period, resulting in DMSPE combining weights near $1/N$. Note that this is in contrast to methods where the combining weights depend on in-sample fit, as discussed in Section 1.2 above.

[16]To make allowance for the fact that data for I/K are released after the end of a quarter, we also computed results using $x_{i,t-1}$ in (1) for I/K, matching the treatment of inflation (see Section 2.1 above). The results reported in Table 1 are qualitatively unchanged.

[17]Results for the other combining methods are similar and available upon request.

[18]Theil (1971) derives the following MSPE decomposition: $\text{MSPE} = (\bar{\hat{r}} - \bar{r})^2 + (\sigma_{\hat{r}} - \rho_{\hat{r},r}\sigma_r)^2 + (1-\rho_{r,\hat{r}}^2)\sigma_r^2$, where $\bar{r}$ ($\bar{\hat{r}}$) is the mean of the actual (predicted) values, σ_r ($\sigma_{\hat{r}}$) is the standard deviation of the actual (predicted) values, and $\rho_{r,\hat{r}}$ is the correlation coefficient between the actual and predicted values, each computed over the forecast evaluation period. Because the actual equity premium is inherently difficult to predict, the actual and predicted values will be weakly correlated. With $\rho_{r,\hat{r}}$ near zero, $\text{MSPE} \approx (\bar{\hat{r}} - \bar{r})^2 + \sigma_{\hat{r}}^2 + \sigma_r^2$. In this case, it is desirable to reduce the forecast variance, $\sigma_{\hat{r}}^2$, as long as it does not lead to a sizable increase in the magnitude of the forecast bias, $\bar{\hat{r}} - \bar{r}$.

[19]We thank John Cochrane and the referee for the analytical insights provided below.

[20]We also generated out-of-sample forecasts for a multiple regression model that only includes the three significant predictors in Panel A of Table 1 (D/Y, D/P, and I/K). This allows us to examine whether a multiple regression model with a relatively small number of predictors reduces the potential near-singularity problem for $\tilde{X}'\tilde{X}$. This model has an R_{OS}^2 statistic of -1.28% over the 1965:1–2005:4 period, however, underperforming the combination forecasts.

[21]For the situation with exogenous regressors and a stable data-generating process, Huang and Lee (2009) show analytically that combination forecasts can outperform kitchen sink model forecasts in finite samples under certain conditions. We emphasize the role of instabilities in the data-generating process in explaining the advantages of forecast combination.

[22]The variance of the kitchen sink model forecast is more than twice that of the largest individual prediction regression model forecast.

[23]Also see the recent survey by Stock and Watson (2006).

[24]Their forecasting model is selected on the basis of an extensive search across potential specifications performed over the entire sample period, so the out-of-sample exercise is not yet a real-time exercise. Indeed, Ludvigson and Ng (2007, p. 181) state that they "are not interested in real-time forecasting per se, but rather in an accurate estimate of the population risk-return relation." In contrast, the present paper focuses on real-time forecasting.

[25]The well-known theoretical model of Campbell and Cochrane (1999) with habit in the utility function generates time-varying risk aversion and equity premium predictability.

[26]The highest value occurs in 1987:3, clearly corresponding to the brief market crash in the fall of 1987.

[27]The kitchen sink model forecast in Figure 5 also implies an implausible equity risk premium for numerous periods.

[28]The data for the macroeconomic variables were downloaded from Federal Reserve Economic Data at http://research.stlouisfed.org/fred2/.

[29]We also computed correlations between the residuals of autoregressive models for combination forecasts of the equity premium and the three macroeconomic variable growth rates. The correlations remain positive, though smaller.

[30]In an investigation of the link between the Fama and French (1993) factors and the real economy, Liew and Vasslou (2000) compute returns for size, value, and momentum portfolios during good and bad periods, where good and bad periods are determined by real GDP growth.

[31]The results are qualitatively similar when we allow for additional lags of y_t and $x_{i,t}$ in (25).

[32]The results for the 2000:1–2005:4 out-of-sample period are qualitatively similar and are available upon request.

[33]Viceira (1997) is the first paper to analyze testing for structural breaks in equity premium predictive regression models. He tests for a structural break in an equity premium predictive regression model based on the dividend-price ratio and monthly data for 1926–1995 and fails to find significant evidence of a break. Differences in results in Viceira (1997) and Paye and Timmermann (2006) and Rapach and Wohar (2006a) can be explained by the latter two studies' use of more recently developed structural break tests.

[34]The $WDmax$ statistic is computed for a pre-specified significance level, and we use the 10% level.

[35]The sequences of $F(l+1|l)$ statistics and ζ_i and λ_i estimates are not reported to conserve space; they are available upon request.

[36]Pástor and Stambaugh (2001) and Kim, Morley, and Nelson (2005) find sizable changes in the unconditional expected equity premium during the prewar period, while it is considerably more stable during the postwar era. This suggests that instabilites in relationships describing the conditional expected equity premium over the postwar period (corresponding to our sample) are primarily due to changes in the sensitivity of the expected equity premium to economic variables rather than substantial changes in the unconditional expected equity premium.

[37]Results are similar when we replace y_{t+1} with real profit or real net cash flow growth in (26) and are omitted for brevity. Pesaran, Pettenuzzo, and Timmermann (2006) recently develop a Bayesian approach to forecasting in the presence of structural breaks. While, as mentioned earlier, Bayesian techniques are beyond the scope of the present

paper, it would be interesting in future research to compare this and other Bayesian techniques with the approaches of this paper in out-of-sample equity premium forecasting.

Table 1
Equity premium out-of-sample forecasting results for individual forecasts and combining methods

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual predictive regression model forecasts						Combination forecasts		
Predictor	R^2_{OS} (%)	Δ (%)	Predictor	R^2_{OS} (%)	Δ (%)	Combining method	R^2_{OS} (%)	Δ (%)
A. 1965:1–2005:4 out-of-sample period								
D/P	$0.34^{\dagger}$	0.55	LTY	−3.09	2.29	Mean	3.58^{**}	2.34
D/Y	$0.25^{\dagger}$	1.41	LTR	0.33	1.30	Median	3.04^{**}	1.03
E/P	0.36	0.64	TMS	−2.96	5.14	Trimmed mean	3.51^{**}	2.11
D/E	−1.42	0.58	DFY	−2.72	−0.83	DMSPE, $\theta = 1.0$	3.54^{**}	2.41
$SVAR$	−12.97	0.13	DFR	−1.10	0.57	DMSPE, $\theta = 0.9$	3.49^{**}	2.59
B/M	−2.60	−0.58	$INFL$	−0.84	1.39			
$NTIS$	−0.91	0.08	I/K	1.44^{*}	2.80	Mean, CT	3.23^{**}	1.25
TBL	−2.78	2.60						
B. 1976:1–2005:4 out-of-sample period								
D/P	−5.08	−0.70	LTY	−5.59	−0.89	Mean	$1.19^{\dagger}$	0.57
D/Y	−6.22	−0.54	LTR	−0.27	1.43	Median	1.51^{*}	0.53
E/P	−1.70	0.75	TMS	−7.24	2.08	Trimmed mean	$1.23^{\dagger}$	0.59
D/E	−2.26	−1.65	DFY	−2.48	−1.18	DMSPE, $\theta = 1.0$	$1.11^{\dagger}$	0.54
$SVAR$	−22.47	0.06	DFR	−2.14	−0.64	DMSPE, $\theta = 0.9$	$1.01^{\dagger}$	0.46
B/M	−4.72	−1.27	$INFL$	−0.08	0.45			
$NTIS$	0.10	0.60	I/K	−3.47	−0.85	Mean, CT	$1.20^{\dagger}$	0.55
TBL	−7.31	−0.82						
C. 2000:1–2005:4 out-of-sample period								
D/P	$10.32^{\dagger}$	12.96	LTY	−0.32	0.24	Mean	3.04^{*}	2.31
D/Y	$10.40^{\dagger}$	12.98	LTR	−1.72	2.57	Median	$1.56^{\dagger}$	0.28
E/P	$8.02^{\dagger}$	9.53	TMS	−4.98	4.23	Trimmed mean	2.98^{*}	2.12
D/E	0.56	0.50	DFY	−0.53	−1.52	DMSPE, $\theta = 1.0$	2.56^{*}	1.65
$SVAR$	−5.62	−1.64	DFR	−2.10	1.76	DMSPE, $\theta = 0.9$	2.66^{*}	1.97
B/M	2.32	3.09	$INFL$	−1.42	0.57			
$NTIS$	−4.09	1.33	I/K	8.96^{*}	9.13	Mean, CT	2.43^{*}	1.32
TBL	−2.50	−0.20						

R^2_{OS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences and risk aversion coefficient of three would be willing to pay to have access to the forecasting model given in column (1), (4), or (7) relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between zero and 1.5 (inclusive). Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) out-of-sample *MSPE-adjusted* statistic; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model given in column (1), (4), or (7) has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 2
Forecast encompassing test results, *MHLN* statistic *p*-values, 1965:1–2005:4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	D/P	D/Y	E/P	D/E	$SVAR$	B/M	$NTIS$	TBL	LTY	LTR	TMS	DFY	DFR	$INFL$	I/K	Mean	Med.	TM	D(1.0)	D(0.9)
D/P		0.32	0.23	0.03	0.13	0.03	0.02	0.00	0.01	0.03	0.00	0.01	0.02	0.04	0.03	0.38	0.33	0.41	0.38	0.36
D/Y	0.38		0.21	0.03	0.12	0.04	0.02	0.00	0.01	0.03	0.00	0.01	0.02	0.04	0.03	0.33	0.28	0.36	0.32	0.31
E/P	0.23	0.19		0.04	0.15	0.02	0.06	0.01	0.01	0.09	0.00	0.03	0.06	0.10	0.05	0.63	0.61	0.66	0.62	0.59
D/E	0.08	0.07	0.14		0.15	0.04	0.10	0.02	0.04	0.20	0.01	0.07	0.14	0.31	0.26	0.85	0.76	0.85	0.84	0.85
$SVAR$	0.01	0.02	0.02	0.02		0.00	0.01	0.00	0.01	0.02	0.00	0.00	0.01	0.03	0.01	0.26	0.11	0.14	0.26	0.24
B/M	0.52	0.41	0.81	0.11	0.17		0.16	0.01	0.02	0.22	0.02	0.16	0.19	0.23	0.12	0.79	0.82	0.81	0.78	0.75
$NTIS$	0.07	0.07	0.20	0.06	0.16	0.03		0.01	0.01	0.22	0.01	0.06	0.11	0.20	0.06	0.76	0.74	0.76	0.74	0.73
TBL	0.01	0.01	0.02	0.04	0.08	0.01	0.01		0.15	0.04	0.01	0.01	0.02	0.04	0.22	0.20	0.13	0.18	0.20	0.20
LTY	0.04	0.04	0.07	0.11	0.10	0.02	0.05	0.24		0.08	0.00	0.02	0.05	0.11	0.23	0.38	0.29	0.37	0.38	0.39
LTR	0.03	0.03	0.07	0.07	0.14	0.01	0.06	0.01	0.02		0.01	0.02	0.04	0.12	0.06	0.52	0.45	0.51	0.51	0.50
TMS	0.01	0.01	0.02	0.02	0.10	0.02	0.02	0.01	0.01	0.06		0.02	0.03	0.04	0.14	0.25	0.17	0.22	0.25	0.25
DFY	0.18	0.16	0.42	0.17	0.18	0.17	0.26	0.01	0.02	0.37	0.02		0.25	0.32	0.16	0.87	0.89	0.89	0.86	0.86
DFR	0.08	0.06	0.18	0.07	0.15	0.05	0.06	0.01	0.02	0.08	0.01	0.06		0.17	0.08	0.84	0.81	0.85	0.83	0.82
$INFL$	0.09	0.07	0.22	0.12	0.16	0.06	0.18	0.01	0.03	0.24	0.01	0.09	0.18		0.16	0.94	0.92	0.94	0.93	0.94
I/K	0.01	0.01	0.02	0.01	0.11	0.01	0.01	0.02	0.02	0.02	0.00	0.00	0.01	0.01		0.27	0.16	0.25	0.27	0.26
Mean	0.01	0.01	0.02	0.01	0.13	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.01	0.02		0.16	0.24	0.31	0.37
Med.	0.02	0.02	0.03	0.01	0.14	0.00	0.00	0.01	0.01	0.04	0.00	0.00	0.01	0.01	0.03	0.63		0.68	0.59	0.55
TM	0.01	0.01	0.02	0.01	0.14	0.00	0.00	0.01	0.01	0.02	0.00	0.00	0.01	0.01	0.02	0.48	0.19		0.41	0.39
D(1.0)	0.01	0.01	0.02	0.00	0.13	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.01	0.02	0.67	0.18	0.32		0.40
D(0.9)	0.01	0.01	0.03	0.00	0.13	0.01	0.00	0.00	0.01	0.03	0.00	0.00	0.01	0.01	0.02	0.59	0.23	0.43	0.56	

The table reports p-values for the Harvey, Leybourne, and Newbold (1998) *MHLN* statistic. The statistic corresponds to a one-sided (upper-tail) test of the null hypothesis that the forecast given in the column heading encompasses the forecast given in the row heading against the alternative hypothesis that the forecast given in the column heading does not encompass the forecast given in the row heading. The table uses the following abbreviations for the combination forecasts: Med. = Median; TM = trimmed mean; D(1.0) = DMSPE, $\theta = 1.0$; D(0.9) = DMSPE, $\theta = 0.9$.

Table 3
Correlation matrix for equity premium forecasts based on individual predictive regression models, 1965:1–2005:4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	D/P	D/Y	E/P	D/E	$SVAR$	B/M	$NTIS$	TBL	LTY	LTR	TMS	DFY	DFR	$INFL$	I/K
D/P	1.00	0.95	0.88	−0.43	0.05	0.78	0.06	−0.80	−0.84	−0.06	−0.16	0.37	0.01	−0.53	−0.29
D/Y	0.95	1.00	0.83	−0.39	−0.07	0.73	0.13	−0.71	−0.76	0.04	−0.07	0.45	0.02	−0.51	−0.20
E/P	0.88	0.83	1.00	−0.68	0.04	0.80	0.13	−0.80	−0.78	−0.05	−0.25	0.34	0.00	−0.54	−0.44
D/E	−0.43	−0.39	−0.68	1.00	−0.04	−0.58	−0.06	0.40	0.42	0.12	0.15	−0.03	0.18	0.71	0.30
$SVAR$	0.05	−0.07	0.04	−0.04	1.00	0.03	−0.02	−0.07	−0.08	0.07	−0.02	0.03	0.01	−0.09	0.01
B/M	0.78	0.73	0.80	−0.58	0.03	1.00	−0.19	−0.70	−0.83	−0.04	−0.08	0.50	0.02	−0.55	−0.31
$NTIS$	0.06	0.13	0.13	−0.06	−0.02	−0.19	1.00	−0.02	0.11	0.14	−0.07	−0.04	−0.05	0.10	−0.19
TBL	−0.80	−0.71	−0.80	0.40	−0.07	−0.70	−0.02	1.00	0.88	0.15	0.58	−0.36	0.01	0.39	0.52
LTY	−0.84	−0.76	−0.78	0.42	−0.08	−0.83	0.11	0.88	1.00	0.11	0.19	−0.50	−0.03	0.49	0.26
LTR	−0.06	0.04	−0.05	0.12	0.07	−0.04	0.14	0.15	0.11	1.00	0.13	0.16	−0.32	0.05	0.05
TMS	−0.16	−0.07	−0.25	0.15	−0.02	−0.08	−0.07	0.58	0.19	0.13	1.00	0.12	0.12	0.00	0.65
DFY	0.37	0.45	0.34	−0.03	0.03	0.50	−0.04	−0.36	−0.50	0.16	0.12	1.00	0.18	−0.10	0.05
DFR	0.01	0.02	0.00	0.18	0.01	0.02	−0.05	0.01	−0.03	−0.32	0.12	0.18	1.00	0.28	0.06
$INFL$	−0.53	−0.51	−0.54	0.71	−0.09	−0.55	0.10	0.39	0.49	0.05	0.00	−0.10	0.28	1.00	0.01
I/K	−0.29	−0.20	−0.44	0.30	0.01	−0.31	−0.19	0.52	0.26	0.05	0.65	0.05	0.06	0.01	1.00

The table reports correlation coefficients for the individual predictive regression model forecasts given in the row and column headings.

Table 4
Correlations between equity premium forecasts and growth rates in three macroeconomic variables, 1965:1–2005:4

(1)	(2)	(3)	(4)
Combining method	Real GDP growth	Real profit growth	Real net cash flow growth
Mean	0.28^{**}	0.35^{**}	0.34^{**}
Median	0.17^{*}	0.24^{**}	0.23^{**}
Trimmed mean	0.31^{**}	0.36^{**}	0.35^{**}
DMSPE, $\theta = 1.0$	0.28^{**}	0.35^{**}	0.34^{**}
DMSPE, $\theta = 0.9$	0.34^{**}	0.36^{**}	0.36^{**}

The table reports correlation coefficients for the equity premium combination forecast given in the row heading and macroeconomic variable growth rate given in the column heading. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5
R^2_{OS} statistics for out-of-sample equity premium combination forecasts during good, normal, and bad growth periods, 1965:1–2005:4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Forecast horizon: one quarter				Forecast horizon: four quarters			
Combining method	Overall	Good	Normal	Bad	Overall	Good	Normal	Bad
A. Sorting on real GDP growth								
Mean	3.58**	1.82	1.71	6.17**	8.19**	3.07	3.63†	11.58**
Median	3.04**	2.67*	0.39	5.02**	6.99**	12.74**	6.35*	5.23**
Trimmed mean	3.51**	2.25†	1.24	5.94**	8.13**	5.41†	4.01†	10.63**
DMSPE, $\theta = 1.0$	3.54**	1.71	1.56	6.26**	7.87**	2.32	3.15	11.46**
DMSPE, $\theta = 0.9$	3.49**	1.60	1.36	6.33**	5.96**	4.71†	0.27	8.27**
B. Sorting on real profit growth								
Mean	3.58**	2.87†	−1.03	7.94**	8.19**	0.93	4.89†	14.72**
Median	3.04**	2.56*	0.21	5.74**	6.99**	1.14	8.00*	10.18**
Trimmed mean	3.51**	2.85†	−0.67	7.47**	8.13**	1.74	5.83*	13.55**
DMSPE, $\theta = 1.0$	3.54**	2.74†	−1.21	8.08**	7.87**	0.16	4.41	14.78**
DMSPE, $\theta = 0.9$	3.49**	2.51	−1.56	8.40**	5.96**	−4.28	2.00	14.70**
C. Sorting on real net cash flow growth								
Mean	3.58**	5.44*	2.17†	4.63*	8.19**	3.29†	8.81**	11.42**
Median	3.04**	4.12**	1.80*	4.25*	6.99**	4.99**	6.17*	9.48**
Trimmed mean	3.51**	5.01*	2.36*	4.47*	8.13**	4.39*	9.13**	10.04**
DMSPE, $\theta = 1.0$	3.54**	5.51*	2.13†	4.52*	7.87**	2.97†	8.50**	11.09**
DMSPE, $\theta = 0.9$	3.49**	5.88*	1.84†	4.15†	5.96**	0.53	6.66*	9.56**

The table reports the Campbell and Thompson (2008) R^2_{OS} statistic. The R^2_{OS} statistics are computed for the entire 1965:1–2005:4 forecast evaluation period (Overall) and three subperiods corresponding the top third (Good), middle third (Normal), and bottom third (Bad) of observations sorted on the macroeconomic variable given in the panel heading. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) out-of-sample *MSPE-adjusted* statistic; the statistic corresponds to a one-sided test of the null hypothesis that the combination forecast given in column (1) has equal expected square prediction error relative to the historical average benchmark forecast against the alternative hypothesis that the combination forecast has a lower expected square prediction error than the historical average benchmark forecast. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6
Macroeconomic variable out-of-sample forecasting results for individual models and combining methods

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Forecast horizon: one quarter						Forecast horizon: four quarters					
	1965:1–2005:4 out-of-sample period			1976:1-2005:4 out-of-sample period			1965:1–2005:4 out-of-sample period			1976:1–2005:4 out-of-sample period		
Predictor	Real GDP growth	Real profit growth	Real net cash flow growth	Real GDP growth	Real profit growth	Real net cash flow growth	Real GDP growth	Real profit growth	Real net cash flow growth	Real GDP growth	Real profit growth	Real net cash flow growth
A. Individual predictive regression model forecasts												
D/P	−0.95	−1.34	−0.79	−1.63	−1.02	−2.61	−3.98	−4.95	−4.84	−3.68	−7.39	−5.02
D/Y	−1.22	−1.15	−0.83	−0.83	−1.52	−1.11	−4.53	−4.56	−3.86	−5.17	−9.95	−2.85
E/P	−0.29	−1.08	0.62	−2.62	−0.07	−1.62	−3.73	−5.31	−0.55	−4.98	−1.68	−0.19
D/E	−0.93	−1.63	−3.66	−6.27	−2.11	−4.50	−0.70	−9.97	−9.31	−11.45	−4.02	−12.19
$SVAR$	−36.57	−7.35	−38.45	−69.76	−6.15	−58.36	−21.90	−15.86	−31.07	−43.96	−19.36	−37.46
B/M	−1.67	−1.82	−1.41	−6.52	−1.50	−2.20	−9.02	−9.09	−5.46	−8.71	−18.78	−6.01
$NTIS$	−0.56	−2.76	−2.20	0.87†	−2.09	−0.67	−6.38	−8.10	−8.95	−4.55	−12.01	−1.79
TBL	−0.93	−0.39	0.17†	−1.59	−2.32	1.63†	0.03*	−11.35	2.23*	−2.43	−14.67	10.61*
LTY	−2.28	−2.70	−1.80	−2.59	−1.70	−0.58	−10.06	−15.68	−9.80	−11.60	−8.78	−1.97
LTR	−19.75	−12.25	−11.18	−19.70	−1.31	−7.09	0.08	−1.39	−2.65	0.78†	−8.02	−11.40
TMS	−5.34	−1.90	−2.34	−9.74	−12.45	−6.63	−9.16	−10.90	−24.18	−33.00	−40.18	−40.72
DFY	−4.77	−2.98	−3.70	−1.80	−2.82	−2.28	−13.94	−2.97	−17.30	−14.45	−5.13	−21.99
DFR	0.55*	−4.08	−2.11	−3.70	−4.23	−3.88	2.13*	0.81*	4.62**	0.69*	2.62*	4.78**
$INFL$	6.32**	−2.71	3.20*	4.64*	0.46	4.18*	15.00**	−8.83	6.10*	11.09*	−6.54	10.13*
I/K	−10.93	5.34**	−10.02	−0.38	3.58**	0.14†	−38.79	17.17**	−67.10	1.21*	15.96**	2.84*
B. Combination forecasts												
Mean	4.48**	2.65*	3.08*	3.47†	2.02†	1.65	10.08*	7.51**	12.05**	11.52*	7.25**	10.85**
Median	4.63**	0.56	2.86*	4.32†	0.78	2.45†	4.45†	0.00	6.37**	4.90†	0.55	7.45**
Trimmed mean	4.28*	2.49*	3.62**	3.29†	1.78†	2.69*	8.74*	4.88*	10.58**	8.96†	4.78**	10.00**
DMSPE, $\theta = 1.0$	4.52**	2.76*	3.11*	3.62†	2.15†	1.70	11.27*	8.36**	13.01**	13.98*	7.99**	11.56**
DMSPE, $\theta = 0.9$	4.35*	2.95*	2.80*	3.43†	1.77	1.45	11.29*	5.23*	10.59**	11.79†	5.39**	10.01**

The table reports the modified Campbell and Thompson (2008) R^2_{OS} statistic (in percent) comparing forecasts from the competing forecasting model given in the row heading to the AR benchmark forecasting model. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) out-of-sample *MSPE-adjusted* statistic; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model given in the row heading has equal expected square prediction error relative to the AR benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the AR benchmark forecasting model. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7
Bai and Perron (1998) multiple structural break test results for real GDP growth predictive regression models and Chow test results for corresponding equity premium predictive regression models, 1947:3–2005:4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Bai and Perron (1998) statistics		Bai and Perron (1998) break dates						Chow test
Predictor	$UDmax$	$WDmax(10\%)$	1st break	2nd break	3rd break	4th break	5th break	6th break	χ^2-statistic
D/P	15.34*	23.54†	1953:2	1959:2	1966:1	1975:1	1985:3	1999:4	44.87**
D/Y	12.16†	18.71†	1953:2	1966:1	1975:1	1985:3	1999:4	–	44.20**
E/P	70.61**	95.35†	1953:2	1959:2	1975:1	1985:3	–	–	16.41*
D/E	14.10*	15.50†	1984:1	–	–	–	–	–	2.83
$SVAR$	123.89**	136.24†	1955:4	1962:1	1968:2	1982:4	–	–	26.84**
B/M	18.93**	26.26†	1953:2	1974:3	1984:2	–	–	–	13.36*
$NTIS$	15.65*	25.20†	1956:4	1962:4	1970:4	1982:4	1990:1	–	16.32†
TBL	11.17*	19.04†	1958:3	–	–	–	–	–	7.46*
LTY	18.33**	28.09†	1958:2	1982:4	–	–	–	–	6.58
LTR	34.79**	38.25†	1966:1	–	–	–	–	–	5.04†
TMS	18.48**	20.05†	1953:1	1959:4	1966:1	1980:2	–	–	42.55†
DFY	9.59	14.03†	1970:4	–	–	–	–	–	3.17
DFR	16.48**	20.15†	1984:3	–	–	–	–	–	5.28†
$INFL$	15.93*	23.53†	1953:1	1961:1	1975:1	1981:1	1999:4	–	12.90
I/K	22.56**	27.98†	1957:3	1963:2	–	–	–	–	4.96

The table reports Bai and Perron (1998) multiple structural break test results for real GDP growth predictive regression models. Columns (2) and (3) report the $UDmax$ and $WDmax(10\%)$ statistics, respectively, corresponding to a one-sided (upper-tail) test of the null hypothesis of zero breaks against the alternative hypothesis of one to eight breaks. Columns (4)–(9) report the break dates estimated by the Bai and Perron (1998) procedure. Column (10) reports the Chow test χ^2-statistic for corresponding equity premium predictive regression models, where the breaks dates tested in the Chow test are the Bai and Perron (1998) breaks dates for the corresponding real GDP growth predictive regression model. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively.

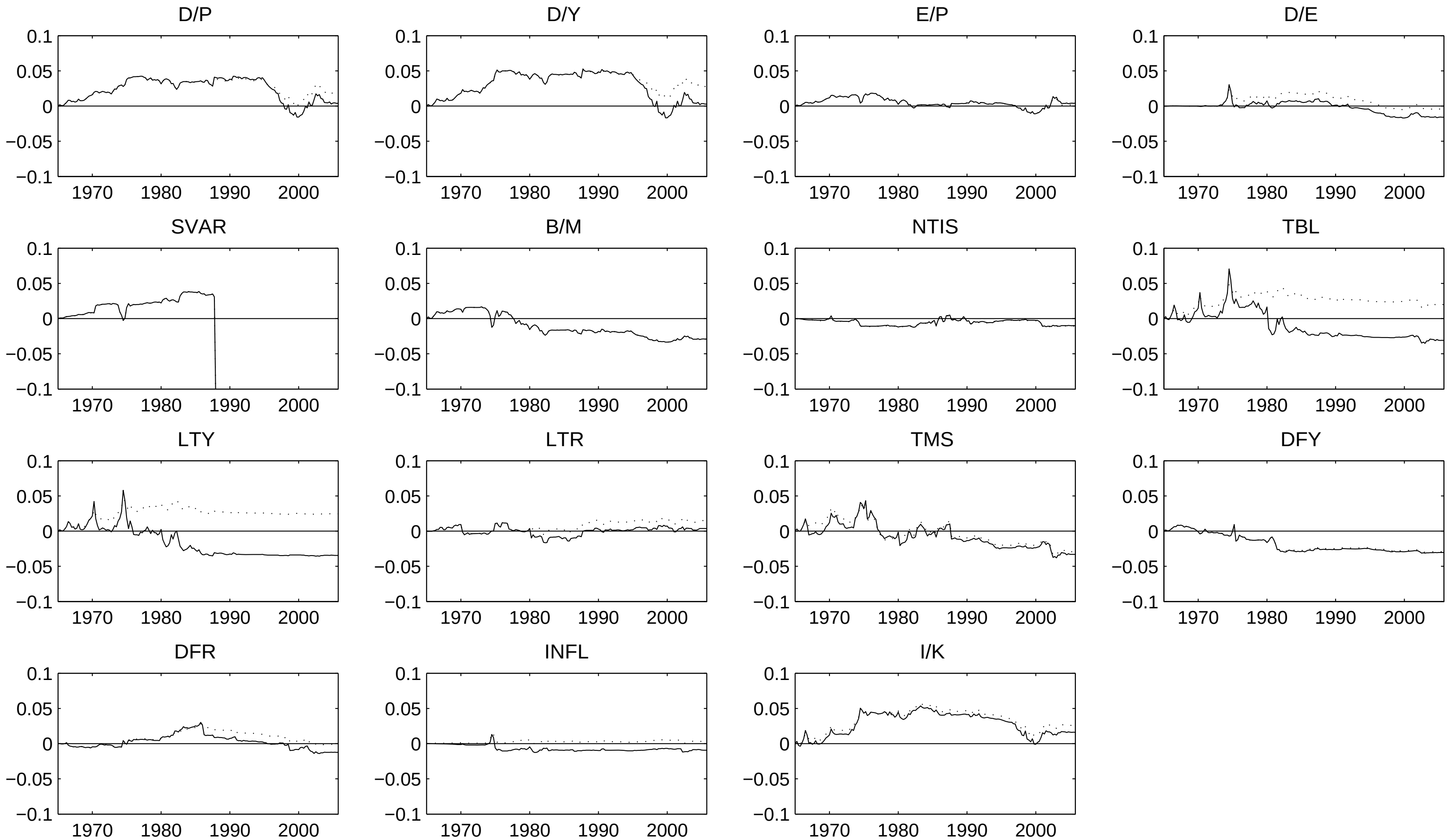

Figure 1

Historical average benchmark forecasting model cumulative square prediction error minus individual predictive regression forecasting model cumulative square prediction error, 1965:1–2005:4

The dotted (solid) line corresponds to individual model forecasts that (do not) impose Campbell and Thompson (2008) restrictions.

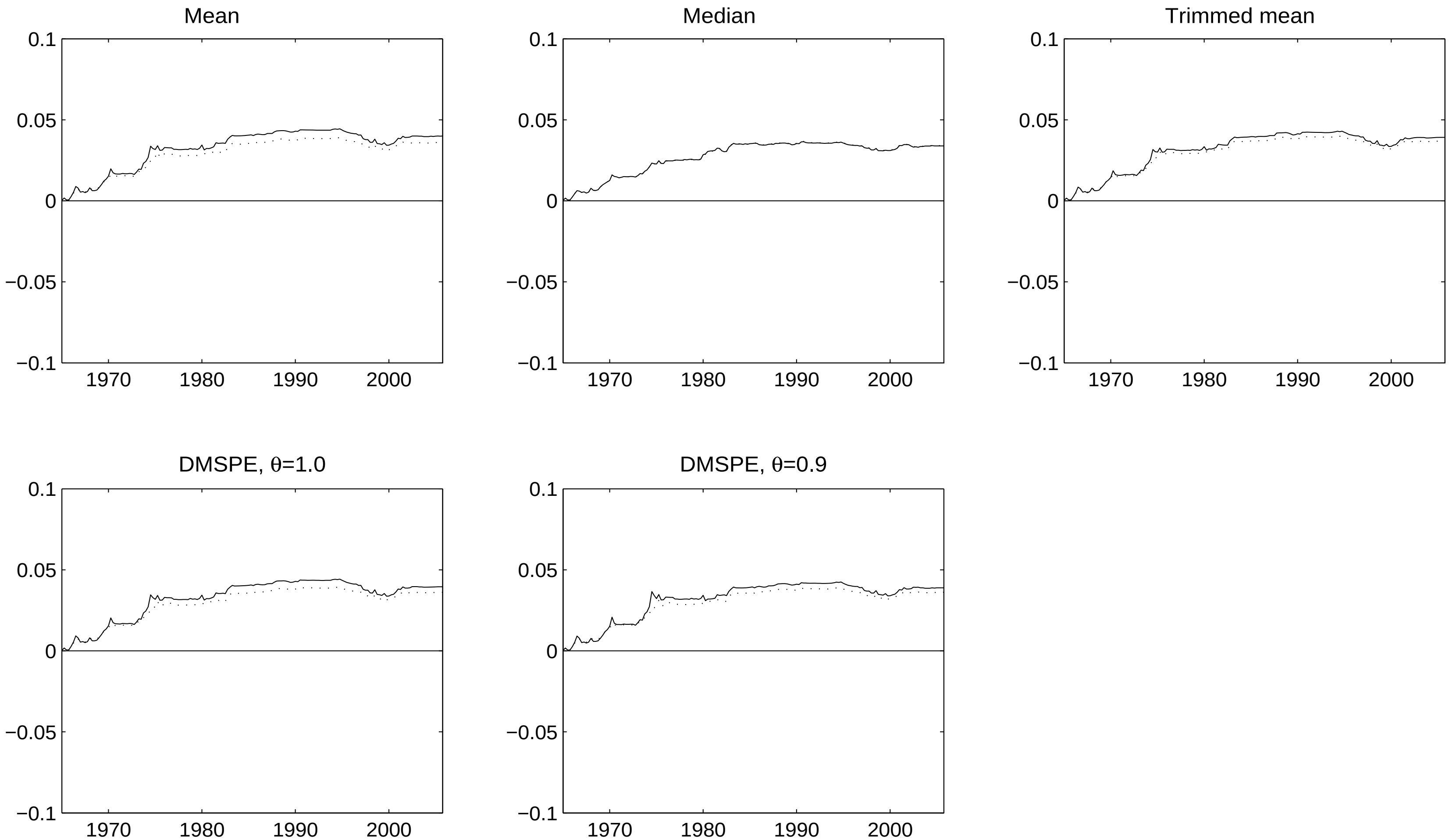

Figure 2

Historical average benchmark forecasting model cumulative square prediction error minus combination forecasting model square prediction error, 1965:1–2005:4

The dotted (solid) line corresponds to combination forecasts based on individual model forecasts that (do not) impose Campbell and Thompson (2008) restrictions.

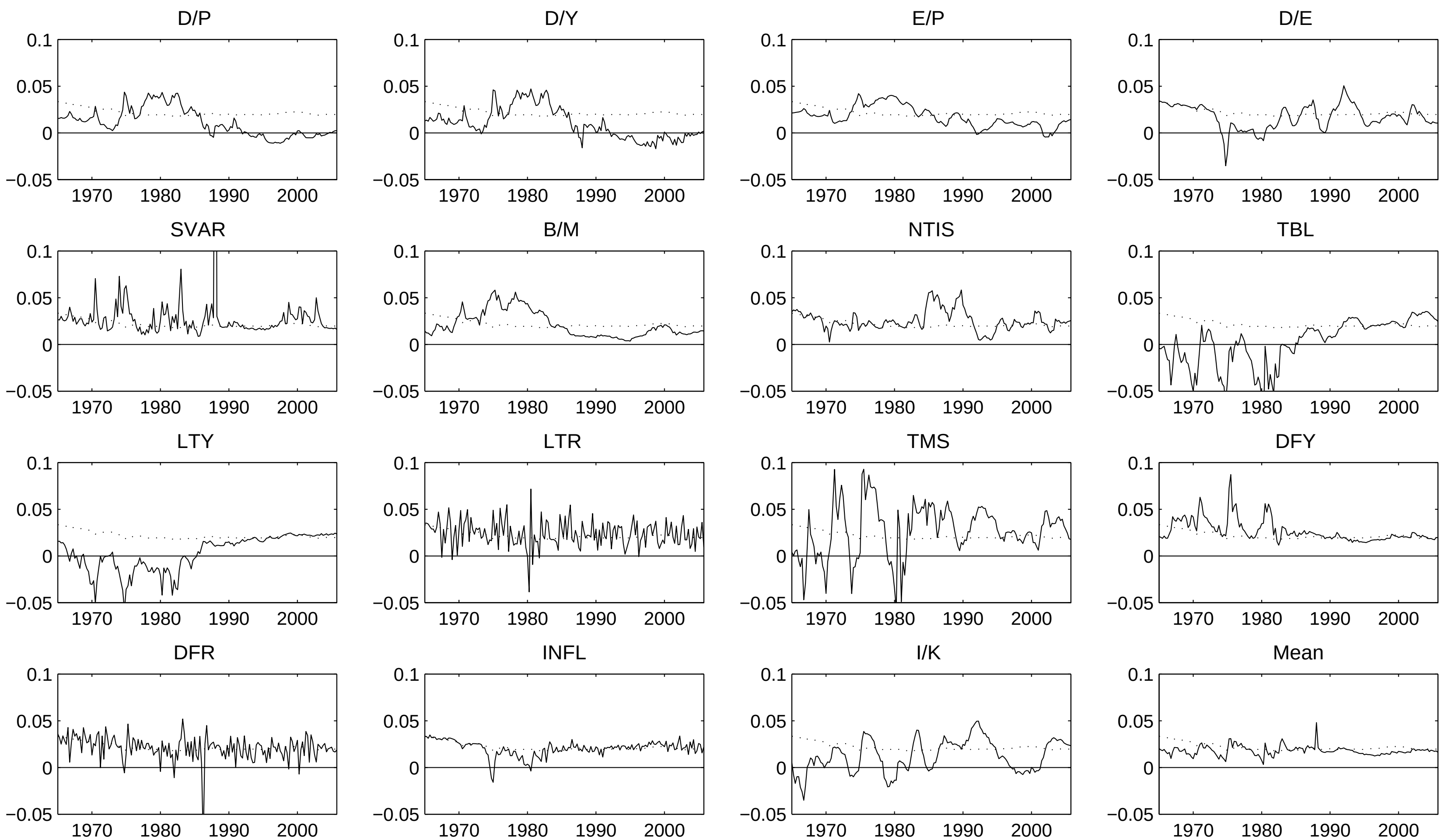

Figure 3
Equity premium forecasts for individual models and mean combining method, 1965:1–2005:4
The solid (dotted) line corresponds to the forecasting model given in the panel heading (historical average forecasting model).

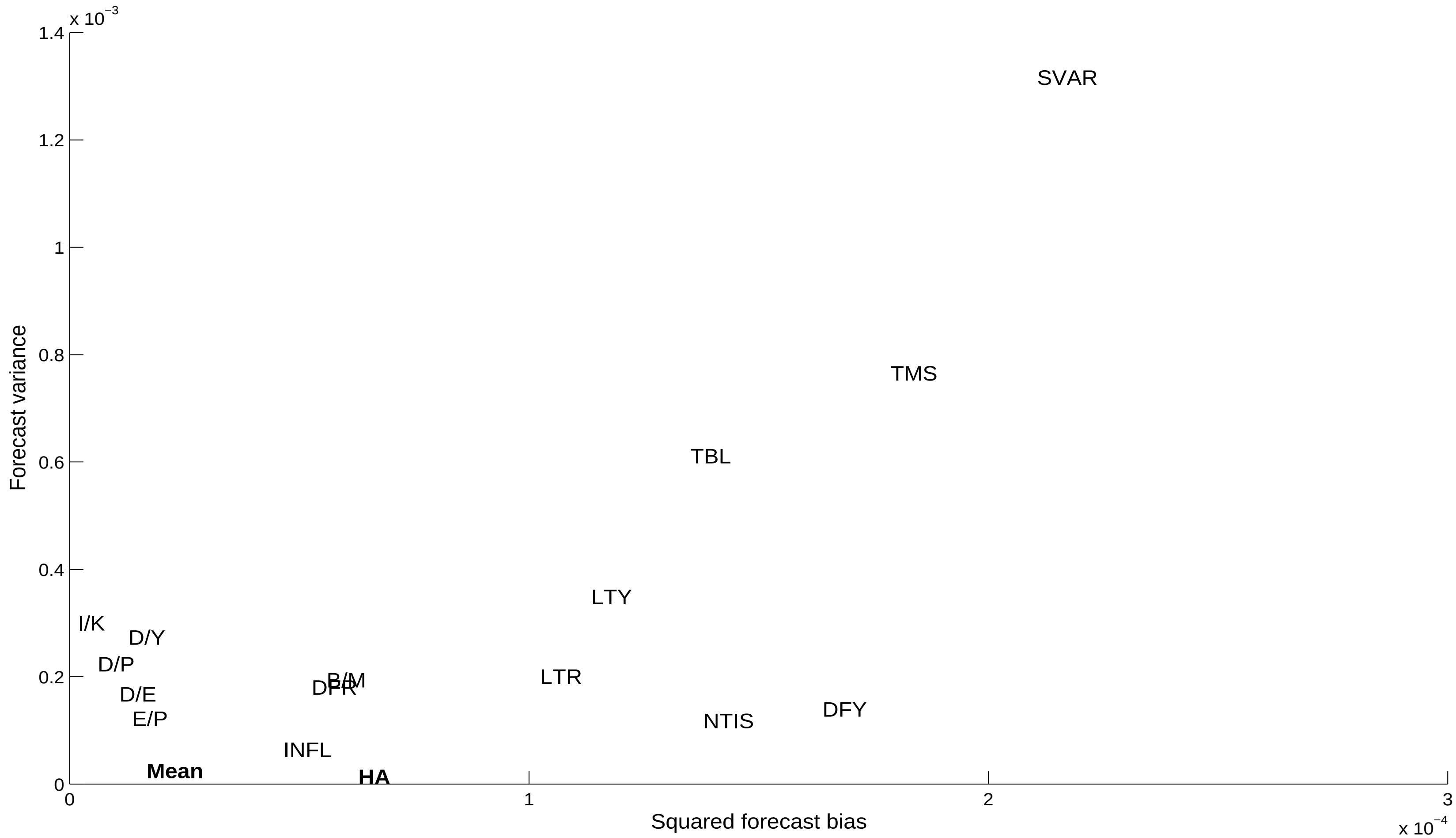

Figure 4
Scatterplot of forecast variances and squared forecast biases, 1965:1–2005:4
HA (Mean) corresponds to the historical average (mean combination) forecast. The other points correspond to the individual predictive regression model forecasts.

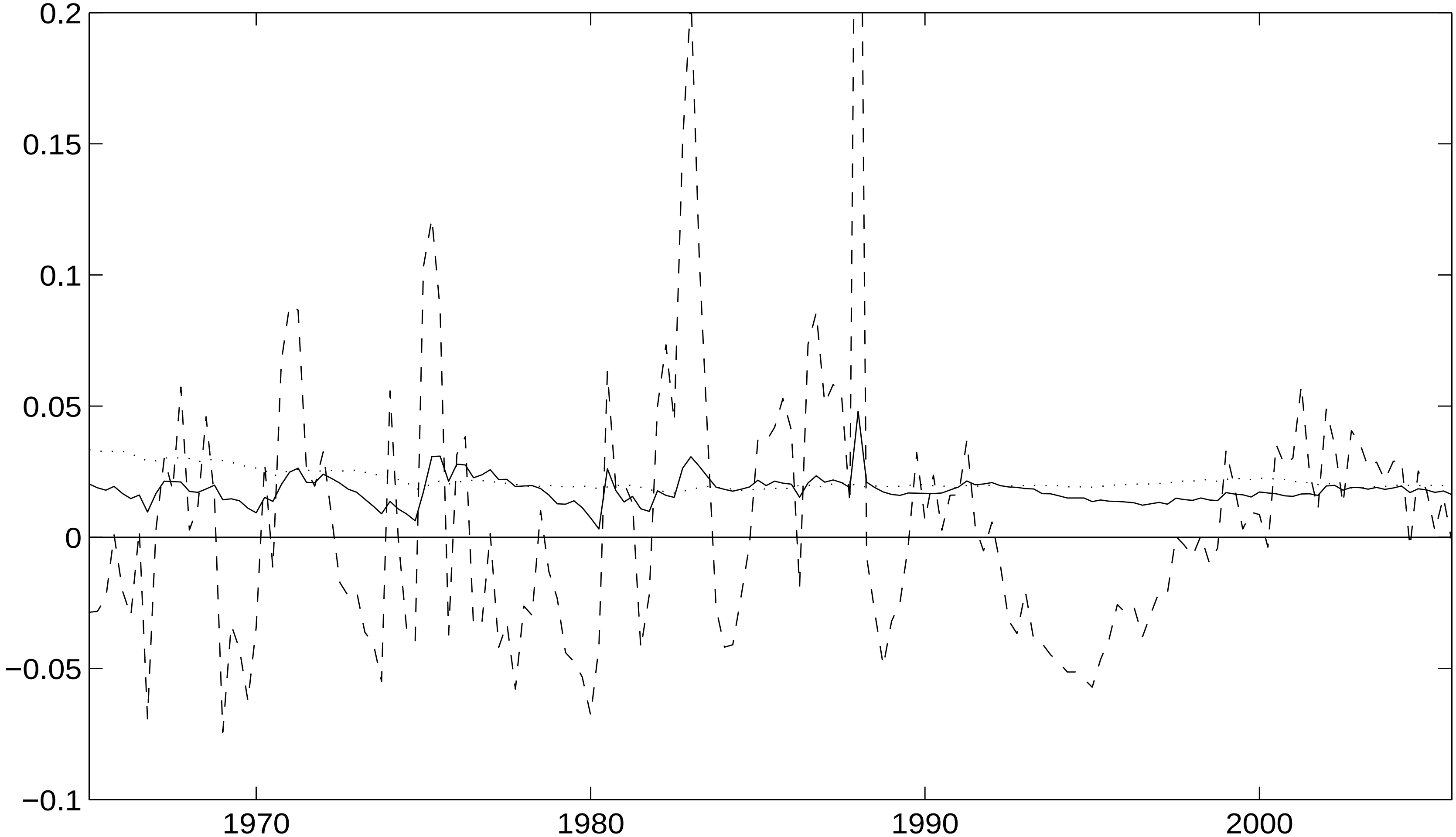

Figure 5

Equity premium forecasts for the mean combining method, historical average, and kitchen sink model, 1965:1–2005:4

The solid (dotted, dashed) line corresponds to the mean combining method (historical average, kitchen sink model) forecast.

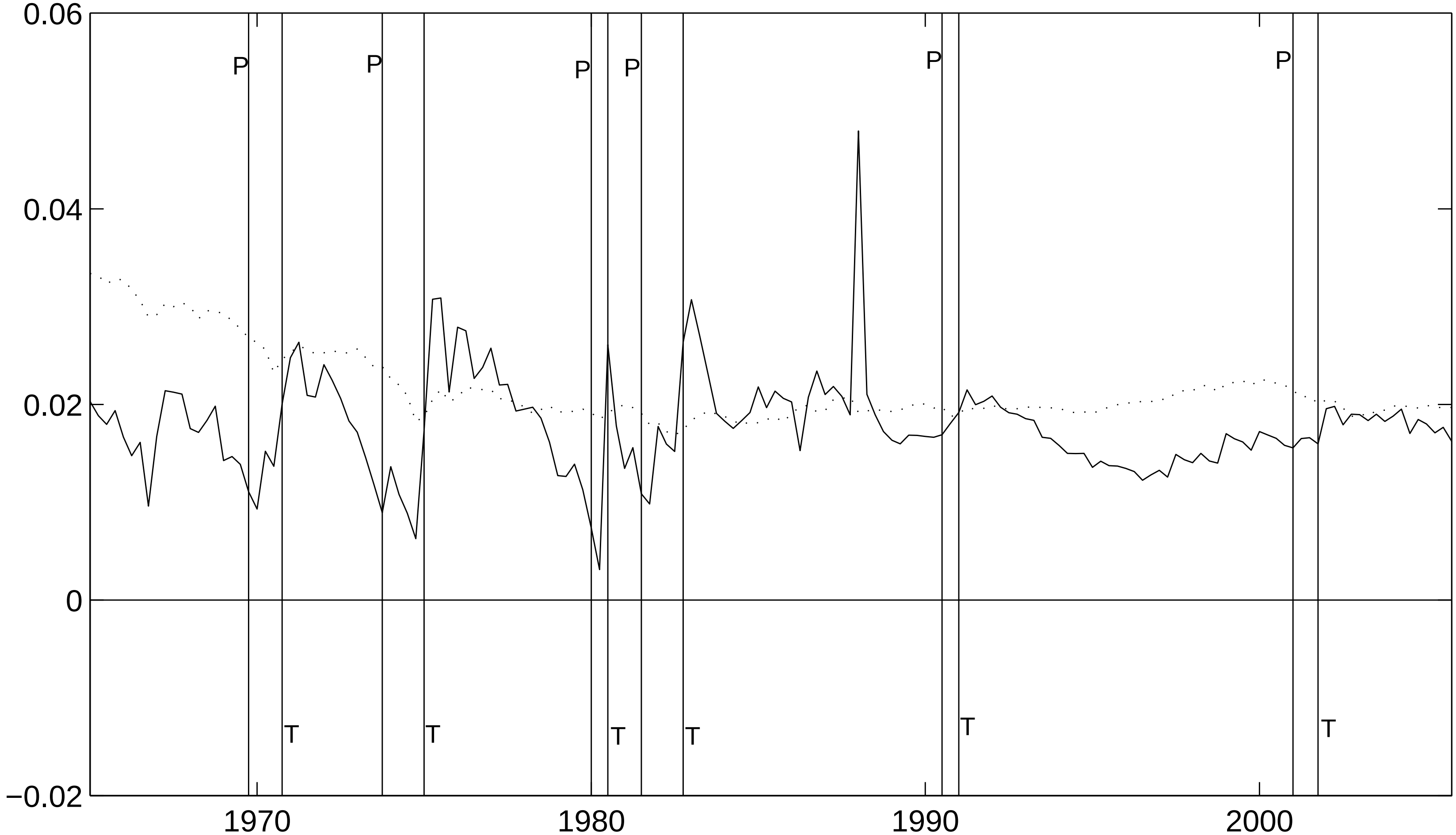

Figure 6
Equity premium forecasts for the mean combining method and NBER-dated business-cycle turning points, 1965:1–2005:4
The solid (dotted) line delineates the mean combination (historical average) forecast. Vertical lines indicate NBER-dated business-cycle peaks (P) and troughs (T).

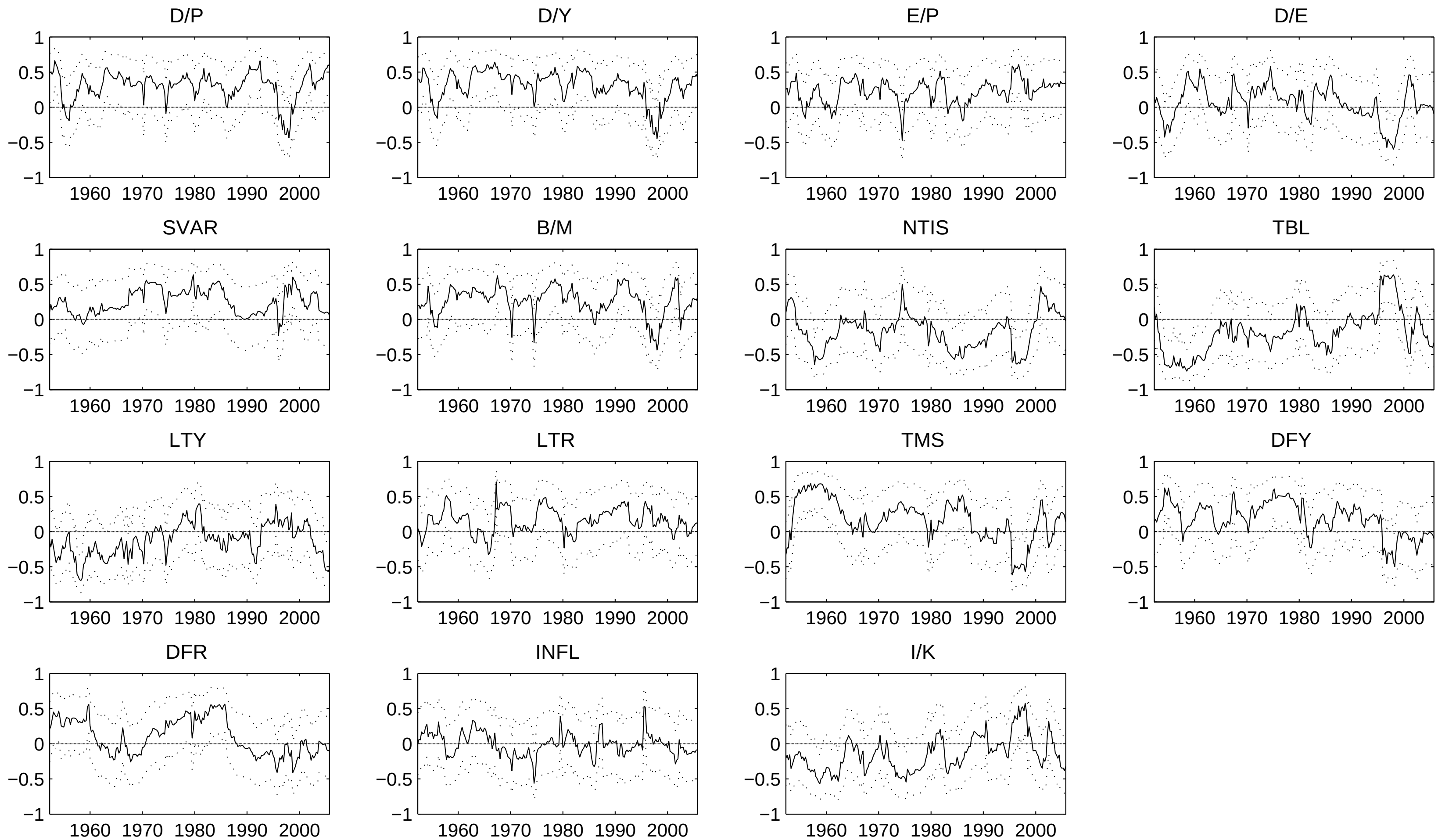

Figure 7

Correlations between the equity premium and individual predictors based on 10-year rolling windows

The date on the horizontal axis gives the end date of the 10-year period. Dotted lines indicate 95% confidence intervals.