mathematics

MDPI

Article

Out-of-Sample Predictability of the Equity Risk Premium

Daniel de Almeida [1,†], Ana-Maria Fuertes [2,†] and Luiz Koodi Hotta [3,*,†]

[1] Department of Statistics, Universidad Carlos III de Madrid, 28903 Getafe, Spain; dani.d.almeida89@gmail.com
[2] Bayes Business School, City University of London, London EC1Y 8TZ, UK; a.fuertes@city.ac.uk
[3] Department of Statistics, Universidade Estadual de Campinas (UNICAMP), Campinas 13083-859, Brazil
* Correspondence: hotta@unicamp.br
† These authors contributed equally to this work.

Abstract: A large set of macroeconomic variables have been suggested as equity risk premium predictors in the literature. Acknowledging the different predictability of the equity premium in expansions and recessions, this paper proposes an approach that combines equity premium forecasts from two-state regression models using an agreement technical indicator as the observable state variable. A comprehensive out-of-sample forecast evaluation exercise based on statistical and economic loss functions demonstrates the superiority of the proposed approach versus combined forecasts from linear models or Markov switching models and forecasts from machine learning methods such as random forests and gradient boosting. The parsimonious state-dependent aspect of risk premium forecasts delivers large improvements in forecast accuracy. The results are robust to sub-period analyses and different investors' risk aversion levels.

Keywords: business cycles; forecast combination; technical indicators; gradient boosting; random forest

MSC: 91B84

JEL Classification: C22; C38; C53; C58; E32; G11; G12; G14; G17

check for updates

Academic Editor: Raymond Lee

Received: 28 November 2024
Revised: 6 January 2025
Accepted: 7 January 2025
Published: 14 January 2025

Citation: de Almeida, D.; Fuertes, A.-M.; Hotta, L.K. Out-of-Sample Predictability of the Equity Risk Premium. *Mathematics* **2025**, *13*, 257. https://doi.org/10.3390/math13020257

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The predictability of the equity market return in excess of the risk-free interest rate plays an important role in several areas of finance, such as asset pricing, asset allocation, and risk management. The prolific literature on equity premium prediction has established that a host of macroeconomic variables have predictive power in the sample for the equity premium; see, for instance [1–4]. Several studies raised skepticism, however, about the predictability of the equity premium upon the empirical finding that the out-of-sample equity premium forecasts from regression models based on macroeconomic indicators are no better than the historical average benchmark; see, e.g., [5–8].

A branch of the empirical finance literature has adduced evidence of out-of-sample equity premium predictability using additional predictors over and above macroeconomic variables, such as technical indicators [9], commodity risk factors [10], liquidity and uncertainty predictors [11], or alternative modelling approaches [12–20]. For a survey of the equity premium predictability literature using predictors, see [21], which examines 26 papers published after [8]. This large number of studies highlights the relevance of the topic.

The present paper is concerned with forecast combination, and within this framework we propose an parsimonious non-linear (state-dependent) approach to predict month-

ahead the equity premium. It proceeds in two stages. At the first stage, we construct equity premium forecasts using different two-state regression models using one, two, or three macroeconomic predictors only. The state variable in these models is observable and novel as defined by an agreement of several technical indicators. Using NBER recessions and expansions data, we underpin this choice of state variable by showing that (binary) technical indicators constructed from equity market prices and/or volume data up to month t are good real-time indicators of the business cycle (expansion versus recession) on month $t+1$. The NBER business cycle indicator is precluded as observable state variable in our models since it is released with a lag because the NBER-dating committee waits long enough so that the existence of a peak or trough is not in doubt (see http://www.nber.org/cycles/, accessed on 6 January 2019). For instance, the Committee's determination of the peak date in December 2007 occurred 11 months after that date and that of the trough date of June 2009 occurred 15 months after that date. The goal is to accommodate expansion-versus-recession variation in the equity premium predictability in a parsimonious manner that exploits the information content of technical indicators. At stage two, we combine the forecasts from the individual two-state regression models using equal weights. We compare these equity premium predictions with the likewise equal-weighted forecasts from *linear* regression models and from conventional Markov-switching models where the state variable is latent. We also consider diffusion indices, an alternative way of conditioning the equity premium forecasts on multiple macroeconomic variables by extracting a few principal components and using them as predictors in a regression. Furthermore, we acknowledge non-linearity in the predictability of the equity premium in a non-parametric manner by also generating forecasts from two machine learning methods: random forests and gradient boosting.

Our paper relates to a theoretical literature that argues that the equity risk premium predictability is not constant over time. The intertemporal capital asset pricing model (ICAPM) of [22] implies that cyclical risk aversion may induce time variation in the degree of equity premium predictability. In line with rational asset pricing, the level of investors' risk aversion may differ during economic expansions and recessions [23] that, in turn, indicates that the predictive nexus between macroeconomic variables and the equity premium may be regime-sensitive.

Our findings speak to an empirical literature that suggests that there is time variation in the equity risk premium predictability related to macroeconomic risk. Ref. [24] finds that relative risk aversion coefficients are not constant over time but differ notably over expansions and recessions. Various empirical studies document that equity returns follow a two-regime process; see, e.g., [3,4,25–32]. In the Bayesian framework proposed by [33] for portfolio decision-making, the certainty-equivalent losses associated with ignoring regime switching from stock market upturns to downturns and vice versa are shown generally to exceed 2% per annum. Refs. [9,34–36] finds that most macroeconomic indicators emit stronger (more informative) signals about future equity premia in recessions than in expansions. Ref. [37] finds that industrial metal returns are positively (negatively) related with future equity premia in recessions (expansions). However, the models employed in the equity premium predictability literature do not explicitly account for state dependence. The exceptions are a few studies that forecast the equity premium with Markov-switching models [34,35,38,39]. Our paper advocates an approach to predict the equity premium through novel, parsimonious threshold models where the state variable is observable as dictated by an agreement of technical indicators.

Our study strengthens the equity premium prediction literature that advocates the combination of forecasts from parsimonious models as an efficient and effective way to blend the information from many macroeconomic predictors. Large-dimensional "kitchen sink" regression models can encounter heightened parameter estimation uncertainty due

to near-collinear predictors and degrees-of-freedom constraints from a low signal-to-noise ratio. This problem can be circumvented through the equal-weights (EW) combination of forecasts from single-predictor models (i.e., an individual macroeconomic predictor per model), which is the approach advocated by [13] and later extended by [40] to the combination of the complete-subset k-predictor regressions (for a given set of K predictors, the complete-subset regressions are all possible regressions with $k \leq K$ predictors). As argued by [13], the data-generating process of expected equity returns is highly uncertain, complex, and constantly evolving; in this context, forecast combinations notably reduce the uncertainty/parameter instability risk associated with reliance on a single model. The equity premium forecast combination approaches deployed by [13,15] employ both EW and weights based on the past forecasting performance of the individual models. Their findings strongly endorse the EW combination. Close alternatives to the EW combination of N forecasts are the median combination, where the forecast is $median\{\hat{r}_{i,t+1}\}$, $i = 1, \ldots, N$, and the trimmed mean combination, which imposes $\omega_{i,t} = 0$ on the p smallest and p largest forecasts and uses weights $\omega_{i,t} = 1/(N - 2p)$ on the remaining individual forecasts. The latter is, strictly speaking, not parameter-free as it requires the choice of p. Ref. [13] shows that the EW combined forecasts are not outperformed by the median and trimmed-mean combined forecasts either.

As a main contribution, considering the time variation in the equity risk premium, we propose the use of two-state models to generate individual forecasts, which are subsequently combined using EW combination to produce the final forecast. We introduce two models: the first is a two-state regression model with an observed state variable determined by technical indicators, while the second employs an unobserved state that follows a Markov-switching process. The out-of-sample (OOS) forecast accuracy of the proposed method is compared to existing equity premium forecasting strategies that combine simple linear models using EW. The evaluation of equity premium forecasts is based on both statistical and economic criteria, such as the asset-allocation problem faced by an investor. Notably, the model with the observed state variable demonstrated the best performance.

Section 2 presents the macroeconomic variables and technical indicators employed in our models as candidate predictors and state variables, respectively. Section 3 discusses the forecasting methods. Section 4 outlines the statistical and economic forecast evaluation criteria. Section 5 discusses the S&P 500 equity premium predictability findings, and Section 6 concludes.

2. Data and Predictors

This section presents the predictors used in the paper. Section 2.1 presents the macroeconomic predictors, and Section 2.2 presents the technical indicators.

2.1. Macroeconomic Predictors

Macroeconomic variables are theoretically motivated as equity premium predictors. The fundamental insight of the ICAPM theory [22] is that, in solving their lifetime consumption decisions under uncertainty, long-term investors care not only about the current level of their invested wealth but also about the future returns on that wealth. More formally, the ICAPM theory contends that, in equilibrium, the expected excess return on an asset is driven by its covariance with current returns on total invested wealth and with (macroeconomic) state variables that contain information about future returns on invested wealth. The state of the economy is the key driver of time-varying expected stock returns; namely, heightened risk aversion during economic downturns requires a higher risk premium and generates equity premium predictability [23,41].

The macroeconomic variables considered as equity premium predictors are taken from [9], which are almost the same used by [8,13]. A more detailed description of each variable can be found in [8]. The only difference among the variables in the papers lies in the definition of volatility. Following [9], we adopt the approach of [42] and define volatility as a function of the absolute value of returns, rather than the square of returns, as it provides a more robust measure. All the stock prices, earnings, and dividends were paid on the S&P 500 index. The variables are as follows.

(1) *Dividend yield* (*DY*): log of the sum of dividends from the last 12 months minus the 1-month lagged stock prices;
(2) *Earnings–price ratio* (*EP*): log of the last 12 months sum of earnings on the S&P 500 Index minus the log of stock prices;
(3) *Dividend payout ratio* (*DE*): log of past 12-month sum of dividends minus the log of sum of past 12 months earnings;
(4) *Equity risk premium volatility* (*RVOL*): [42] estimator $RVOL_t = [\sqrt{6\pi}(|r_t| + \ldots + |r_{t-11}|)]/12$;
(5) *Book-to-market ratio* (*BM*): ratio of book-to-market value to market value for the Dow Jones Industrial Average;
(6) *Net equity expansion* (*NTIS*): ratio of past 12-month sum of net issues by NYSE listed stocks to end-of-year market capitalisation;
(7) *Treasury bill* (*TBL*): interest rate on a three-month U.S. Treasury bill (secondary market);
(8) *Long-term rate of returns* (*LTR*): return on long-term U.S. government bonds;
(9) *Term Spread* (*TMS*): long-term government bond yield minus Treasury bill rate;
(10) *Default yield spread* (*DFY*): difference in Moody's BAA- and AAA-rated corporate bond yields;
(11) *Default return spread* (*DFR*): long-term corporate bond return minus long-term government bond return;
(12) *Inflation* (*INFL*): change in the logarithm of the Consumer Price Index (CPI). A one-month lag is used to account for the delay in the release of the index.

2.2. Technical Indicators

Technical indicators are binary variables constructed primarily from historical equity market prices. In a recent paper, Ref. [9] probes that technical indicators have additional predictive information for the equity premium over and above macroeconomic variables. The intuition is that technical indicators are able to signal changes in investor sentiment, which are known to correlate with future stock returns, as borne out by the evidence in [9]. Following this wisdom, we entertain the same set of 14 trend-following type technical indicators as [9]. Six of these technical indicators belong to the *moving average* MA(*a*,*b*) family:

$$\mathrm{MA}_t(a,b) = \begin{cases} 1 & \text{if} \quad (\sum_{j=0}^{a-1} P_{t-j})/a \leq (\sum_{i=0}^{b-1} P_{t-i})/b, \\ 0 & \text{if} \quad (\sum_{j=0}^{a-1} P_{t-j})/a > (\sum_{i=0}^{b-1} P_{t-i})/b, \end{cases} \tag{1}$$

where P_t is the month t level of the S&P 500 index; we consider $a = 1, 2, 3$ and $b = 9, 12$ months.

Two other technical indicators also based solely on equity prices belong to the *momentum* family:

$$\mathrm{MOM}_t(l) = \begin{cases} 1 & \text{if} \quad P_t > P_{t-l}, \\ 0 & \text{if} \quad P_t \leq P_{t-l}, \end{cases} \tag{2}$$

with $l = 9, 12$ months. The final six technical indicators belong to a family that uses volume data:

$$\mathrm{VOL}_t(a,b) = \begin{cases} 1 & \text{if} \quad (\sum_{j=0}^{a-1} \mathrm{OBV}_{t-j})/a \leq (\sum_{j=0}^{b-1} \mathrm{OBV}_{t-j})/b, \\ 0 & \text{if} \quad (\sum_{j=0}^{a-1} \mathrm{OBV}_{t-j})/a > (\sum_{j=0}^{b-1} \mathrm{OBV}_{t-j})/b, \end{cases} \tag{3}$$

where $\mathrm{OBV}_t = \sum_{i=1}^{t} \mathrm{VOL}_i \times D_i$ is the *on-balance volume* with Vol_i denoting the volume traded in the S&P 500 index on month i and D_i denoting a directional indicator that takes value 1 if $P_i - P_{i-1} \geq 0$ and -1 if $P_i - P_{i-1} < 0$. As in [9], we compute signals for $a = 1, 2, 3$ and $b = 9, 12$.

We depart from [9] in (a) defining an *agreement technical indicator* that blends all 14 technical indicators, (b) utilising the agreement indicator as an observable state variable in parsimonious threshold models of the equity premium. The pth agreement technical indicator is defined as

$$A_{p,t} = \mathcal{I}(\sum_{i=1}^{14} \mathrm{TECH}_{i,t} \geq p), \tag{4}$$

where $\mathrm{TECH}_{i,t}$ are the technical indicators just presented and $p \in \{1, \ldots, 14\}$, and $\mathcal{I}(condition)$ is equal to one if the condition is met, and zero otherwise.

More recent articles using technical indicators to forecast the equity risk premium, in addition to the many papers listed in [21], include [43,44].

3. Forecasting Methodology

Ref. [13] advocates forecasting the equity premium through an EW combination of forecasts from single-predictor linear models. This approach was later extended by [40] to include the combination of complete-subset k-predictor regressions. In the introduction, we listed various papers suggesting that there is time variation in the equity risk premium. In this context, we propose the use of two-state models to generate single forecasts, which are subsequently combined using an EW approach to produce the final forecast. We introduce two models. The first, presented in Section 3.2, is a two-state regression model where the state variable is observed and depends on the technical indicators discussed in Section 2.2. In the second model, the state is unobserved and follows a Markov-switching process.

For the sake of completeness, we present the linear models, as suggested by [13,40], in Section 3.1, and the combined forecasts in Section 3.4.

Since machine learning methods have become very popular for predicting the equity risk premium (see, for example, [45,46]), in Section 3.5 we present two methods that are used in the comparison.

Let r_{t+1} denote the equity premium defined as the return on a broad stock market index, including dividends, in excess of the risk-free interest rate from month t to month $t+1$. Let $x_{i,t}$ represent the value of the ith macroeconomic predictor at month t, $i = 1, \ldots, N$. At month t, we want to predict r_{t+1} based on the observations from the return and predictors from month 1 to t.

3.1. Linear Model Forecasts

The simplest predictive model for the equity risk premium is the single-predictor linear regression. For the ith macroeconomic predictor, $x_{i,t}$, the ith model is given by:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}, t = 1, \ldots, T, \quad i = 1, \ldots, N, \tag{5}$$

where $\varepsilon_{i,t+1}$ is the error term.

For the ith model, let $E_t^{(i)}(\cdot)$ denote the conditional expectation based on the information available at time t, i.e., $\{(r_j, x_{i,j}, j = 1, \ldots, t\}$. A forecast of the OOS equity premium, $E_t^{(i)}(r_{t+1})$, can be constructed from model (5) as $\hat{r}_{t+1}^{(i)} = \hat{\alpha}_i + \hat{\beta}_i x_{it}$, where $(\hat{\alpha}_i, \hat{\beta}_i)$ are parameter estimates based on data up to time t. By assuming no-predictability or that $x_{i,t}$ has no predictive ability (imposing $\beta_i = 0$), the constant expected equity premium model arises; accordingly, the month-ahead forecast is the historical average (HA) of excess returns, i.e., $\bar{r}_{t+1} = (\sum_{i=1}^{t} r_i)/t$. If $x_{i,t}$ conveys information about the future equity risk premium, then $\hat{r}_{t+1}^{(i)}$ should outperform the no-predictability HA benchmark.

A natural extension of model (5) is the k-predictor LIN model. There are $\binom{N}{k}$ candidate models in the complete-subset LIN (k) family of models, where N is the number of macroeconomic predictors available. The combination number $\binom{N}{k}$ gives the total number of ways to combine N numbers in groups of k. Ref. [40] averages the forecasts across complete-subset regressions, that is, all possible regressions with $k \leq N$ predictors where N is the total number of candidate predictors. Formally, for each pair (i, j) of regressors, the LIN (k=2; two-regressor) model can be written as

$$r_{t+1} = \alpha_{ij} + \beta_i^{(i,j)} x_{i,t} + \beta_j^{(i,j)}] x_{j,t} + \varepsilon_{ij,t+1}, \tag{6}$$

where $i, j = 1, \ldots, N (i \neq j)$. Ref. [40] demonstrates that the quality of the OOS complete-subset combined forecasts can quickly deteriorate as k increases. Given the high dependence among macroeconomic predictors, we consider complete-subset LIN models with $k = \{1, 2, 3\}$.

3.2. Non-Linear Two-State Regression Model Forecasts

Ref. [13,15] shows that simple linear regressions using macroeconomic predictors and equal-weighted (EW) combinations thereof, respectively, produce larger out-of-sample (OOS) forecast accuracy gains versus the historical average benchmark in recessions than in expansions. Other studies showing that the predictability of the equity premium is greater in recessions are [9,15,34–36], inter alia. We consider an extension of model (5) to accommodate a specific type of non-linear predictability, state dependence, where the state variable is defined as an observable.

The two-state regression (TSR) model for the single macroeconomic variable $x_{i,t}$, with $A_{p,t}$ as the state variable, can be written as

$$r_{t+1} = \alpha_{i0}^{(p)} + \alpha_{i1}^{(p)} A_{p,t} + \beta_{i0}^{(p)} x_{i,t} + \beta_{i1}^{(p)} x_{i,t} \mathcal{A}_{p,t} + \varepsilon_{t+1}^{(p)}. \tag{7}$$

where the state variable $A_{p,t}$ is 1 of the 14 agreement technical indicators ($A_{p,t}, p = 1, \ldots 14$) defined in (4). As in the LIN models, we consider the complete-subset TSR models with $k = \{1, 2, 3\}$. Since we have 12 macroeconomic variables we have 12, 66, and 220 models with one, two, and three regressors, respectively. For each estimation window and state-dependent model, we select the agreement indicator that minimises the in-sample mean square adjustment error. In our empirical analysis, we also consider a version of the TSR model where only the coefficients of the explanatory variables depend on the state variable, not the intercept.

3.3. Non-Linear Markov-Switching Model Forecasts

The Markov-switching (MS) extension of the linear predictive model (5) relies on a latent state variable, S_t, that is assumed to follow an order-one Markov chain with transition matrix

$$\Pi = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}, \tag{8}$$

where the element p_{ij} denotes the probability of switching from state i on month t to state j on month $t+1$ [34,35,38,39,47]. We consider the following MS generalization of the LIN ($k = 1$) model:

$$r_{t+1} = \gamma_{iS_t} + \delta_{iS_t} x_{i,t} + \eta_{i,t+1}, \tag{9}$$

where γ_{iS_t} and δ_{iS_t} are the predictive intercept and slope, respectively, at state S_t. Again, we consider two cases: with the equal intercept ($\gamma_{i1} = \gamma_{i2}$) restriction and without it. As with the previous two modelling classes (LIN and non-linear TSR), for the MS class of non-linear models we also generate forecasts from the complete-subset $k = \{1, 2, 3\}$ specifications.

Using the MATLAB code from [48], which we gratefully acknowledge, we estimate the MS models by maximum likelihood using the Hamilton's filter to obtain the unobserved transition matrix Π.

3.4. Combined Forecasts

For each family of models—either LIN regression models, TSR models, or MS models—we have a total number $S \equiv \binom{N}{k}$ of forecasts available for combination (where N is the number of available macroeconomic predictors), which are obtained by recursive estimation of the S complete-subset regressions associated. For instance, $k = 1$ represents the simple (single-predictor) case, $k = 2$ the two-predictor case, and so forth.

Let $\hat{r}_{i,t+1}$ denote the equity premium forecast obtained from the ith model, $i = 1, \ldots, S$. The mean of all S forecasts is usually referred to as the EW forecast combination

$$\hat{r}^{EW}_{t+1} = \sum_{i=1}^{S} \omega_{i,t} \hat{r}_{i,t+1} = \sum_{i=1}^{S} \frac{1}{S} \hat{r}_{i,t+1} = \text{mean}\{\hat{r}_{i,t+1}\}, \tag{10}$$

which represents the simplest (parameter-free) combination approach. We use the notation EW-LIN ($k = j$) to refer to the equity premium prediction obtained as the equal-weighted (EW) combination of S forecasts from linear models with j predictors; similar notation is used for the non-linear model combinations of the preceding two subsections as EW-TSR ($k = j$) and EW-MS ($k = j$), respectively.

3.5. Machine Learning Forecasts

Machine learning methods have often been used to predict the equity risk premium and stock returns. For example, just to cite three papers, Ref. [45] conducts a comparative analysis of machine learning methods to measure equity risk premiums and finds economic gain; Ref. [49] examines asset pricing in the Chinese stock market and finds "high predictability of large stocks and state-owned enterprises over longer horizons"; and [46] compares several machine learning methods and concludes that "the competing forecasting models generally fail to outperform the historical average benchmark". In conclusion, we can say that the efficiency of machine learning methods in predicting the equity risk premium and stock returns, although not yet conclusive, is promising.

We generate equity premium forecasts from two popular ensemble models based on decision trees for regression—gradient boosting (GB) and random forest (RF)—that accommodate general forms of non-linearity. The GB approach, proposed by [50], builds many trees in a gradual, additive, and sequential manner, so that the trees are fitted one at a time, where each new tree helps to correct errors made by the previously trained tree. This procedure is aimed at reducing the forecast bias and variance of a single tree and generalises boosting methods by allowing optimisation of an arbitrary differentiable loss function.

The RF is constructed by fitting multiple decision trees at training time and outputting the mean prediction of the individual trees. The algorithm for random forests applies

the technique of bootstrap aggregating, or bagging, to tree learners. Given a training set $X = x_1, \ldots, x_n$ with responses $Y = y_1, \ldots, y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples. In each of $b = 1, \ldots, B$: iterations (training) the bagging proceeds as follows:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b;
2. Train a modified regression tree f_b on X_b, Y_b, which is a tree learning algorithm that selects, at each candidate split in the learning process, a random subset of the features. This process is sometimes called "feature bagging".

After training, predictions for unseen samples can be made by averaging the predictions from all the individual regression trees on X as $\hat{f} = B^{-1} \sum_{b=1}^{B} f_b(X)$. The bootstrapping procedure improves the forecast performance, i.e., the mean square prediction error by decreasing the forecast variance without increasing the bias. If some features are very strong predictors for the response variable, these features will be selected in several of the B trees, which will become strongly correlated. The motivation for "feature bagging" is to mitigate the correlation of the individual trees in the bootstrap sample. We deploy the RF and GB decision trees with the *sklearn* package from Python. The RF hyper-parameters considered in our application are the number of trees (or n-estimators) set at $100, 200,$ and 400 and the number of features to consider when looking for the best split (max features) at 2 and 4, while the other hyper-parameters are left at the default values. The GB hyper-parameters used are the number of trees (n-estimators) again at $100, 200,$ and 400 and the learning rate at 0.1 and 0.01, while the other hyper-parameters are left at the default values.

4. Out-of-Sample Forecast Evaluation Tools

This section describes the methods used in the paper to evaluate the J one-month-ahead out-of-sample (OOS) forecasts obtained sequentially through expanding estimation windows. Consider that the first estimation window uses the initial T observations, such that the first prediction, denoted as $\hat{r}_{T+1}$, is for r_{T+1}.

4.1. Statistical Evaluation

The first statistical criterion, introduced by [12], measures the reduction in the mean squared error (MSE) of a specific model-based forecasting approach versus the HA benchmark:

$$R^2_{OOS} = 1 - \frac{\sum_{j=1}^{J} (\hat{r}_{T+j} - r_{T+j})^2}{\sum_{j=1}^{J} (\bar{r}_{T+j} - r_{T+j})^2}, \tag{11}$$

where $\hat{r}_{T+j}$ is the OOS forecast for month $T + j$ based on the estimation window starting on month 1 and ending on month $T + j - 1$, and $\bar{r}_{T+j+1}$ is the HA in the same estimation window.

We report the above R^2_{OOS} statistic based on the OOS forecasts associated to all J recursive estimation windows, and two additional measures computed, separately, on the expansionary (Exp) and recessionary (Rec) months according to the NBER-cycle dating. We also report an R^2_{OOS} measure that uses instead the [13] simple average of forecasts from simple LIN (k = 1) models, denoted as EW-LIN (k = 1) in Section 3.1, as the benchmark. Higher R^2_{OOS} indicates better out-of-sample performance relative relative to the benchmark.

Next, we formally test the null hypothesis that the benchmark, HA, or EW-LIN (k = 1), has at least as good predictive content for the month-ahead equity premium as the model-based forecast at hand; i.e., H_0: $R^2_{OOS} \leq 0$ against H_A: $R^2_{OOS} > 0$. For this purpose, we utilise the [51] test, which is based on the MSE-adjusted statistic and can be cast as an extension of the [52,53] tests that allow comparisons between nested models. The results of the significance tests for R^2_{OOS} with the EW-LIN (k = 1) as benchmark should be interpreted

with caution in the context of the EW-MS forecasts since the MS models are non-linear in the parameters.

Because the R^2_{OOS} (and associated tests) is an overall measure of forecast performance based on the mean prediction error, a point statistic, it can mask important instability in forecast performance. To gauge the dynamics of the forecast performance over the OOS months, in the third analysis, we plot the differential cumulative square error (ΔCSE). The ΔCSE at time t, for $t = T+1, \ldots, T+J$, denoted as ΔCSE_t, is defined as

$$\Delta\text{CSE}_t = \sum_{s=T+1}^{t} \left[(\bar{r}_s - r_s)^2 - (\hat{r}_s - r_s)^2 \right], \quad t = T+1, \ldots, T+J, \tag{12}$$

such that a positively (negatively) sloped ΔCSE_t graph indicates that the forecasting model at hand consistently outperforms (underperforms) the HA benchmark; a switch from a positive to a negative slope or vice versa indicates unstable forecast performance. We also define a ΔCSE_t variant based on the EW-LIN ($k = 1$) forecasting approach of [13] as the relevant benchmark.

4.2. Economic Evaluation

We carry out an asset allocation exercise to compare the economic merit of the equity premium forecasts. Our representative investor forms a portfolio at time t by allocating ω_t of her total wealth to stocks and the remainder $(1 - \omega_t)$ to risk-free bills. Accordingly, her total wealth on month $t+1$ is

$$W_{t+1} = [(1 - \omega_t) \exp(r^f_{t+1}) + \omega_t \exp(r^f_{t+1} + r_{t+1})] W_t, \tag{13}$$

where the risk-free interest rate, r^f_{t+1}, in this equation is the 1-month U.S. Treasury bill rate, ensuring that the investor is not exposed to interest rate risk. The expression r^f_{t+1} represents the rate observed at portfolio formation time t; it is not a prediction and is applied in the evaluation exercise as the risk-free rate for the period from t to $t+1$. The equity risk premium, r_{t+1}, and the risk-free interest rate, r^f_{t+1}, are continuously compounded. We assume that the investor maximises the short-term expected 1-month-ahead wealth, which excludes any intertemporal hedging component in the choice of the portfolio weights. Hence, the portfolio weights on month t are the solution to the following optimising problem:

$$\omega^*_t = \arg\max_{\omega_t} E_t[U(W_{t+1})], \tag{14}$$

where the utility function $U(W_{t+1})$ is defined according to the investor's preferences. We consider an investor with mean–variance preferences and corresponding utility $U(W_{t+1}) = E_t[W_{t+1}] - \frac{\gamma}{2} Var_t[W_{t+1}]$, where the parameter γ reflects the investor's absolute risk aversion. In this mean–variance preferences setting, the optimal proportion of wealth allocated to equities on month $t+1$ is

$$\omega^*_t = \frac{\exp(\hat{r}_{t+1} + \hat{\sigma}^2_{t+1}/2) - 1}{\gamma \exp(r^f_{t+1}) \exp(\hat{\sigma}^2_{t+1} - 1) \exp(2\hat{r}_{t+1} + \hat{\sigma}^2_{t+1})}, \tag{15}$$

where $\hat{r}_{t+1}$ is the OOS model-based forecast of the equity premium or predicted mean excess return, and $\hat{\sigma}^2_{t+1}$ is the OOS predicted variance; following [9,12], inter alia, we obtain the latter as the sample variance estimated over the past five-year rolling window of historical monthly excess returns in order to allow for a time-varying variance.

When the equity risk premium prediction is zero, $\hat{r}_{t+1} = 0$, the optimal proportion of wealth allocated to equities on month $t+1$, Equation (15), simplifies to

$$\omega_t^* = \frac{\exp(\hat{\sigma}_t^2/2) - 1}{\gamma \exp\left(r_{t+1}^f\right)[\exp(\hat{\sigma}_t^2) - 1]}.$$

This result shows ω_t^* that becomes inversely proportional to the risk aversion parameter (γ) and the risk-free rate (r_{t+1}^f), and depends on the level of equity market volatility ($\hat{\sigma}_t^2$). If the level of volatility is low, applying a first-order Taylor expansion gives $\exp(\hat{\sigma}_t^2/2) - 1 \approx \hat{\sigma}_t^2/2$ in the numerator, and $\exp(\hat{\sigma}_t^2) - 1 \approx \hat{\sigma}_t^2$ in the denominator. Thus, the optimal weight becomes

$$\omega_t^* = \frac{1}{2\gamma \exp\left(r_{t+1}^f\right)}.$$

We also consider an investor with constant relative risk aversion (CRRA) preferences, or power utility $U(W_{t+1}) = W_{t+1}^{1-\gamma}/(1-\gamma)$, where γ reflects the investor's coefficient of relative risk aversion, in order to parsimoniously capture the higher moments of the portfolio return distribution. In this CRRA preferences setting, the optimal portfolio weight is given by

$$\omega_t^* = \frac{\hat{r}_{t+1} + \hat{\sigma}_{t+1}^2/2}{\gamma \hat{\sigma}_{t+1}^2}, \tag{16}$$

where $\hat{r}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the predicted mean and variance, as explained above.

To provide some intuition, consider the case where the equity premium forecast ($\hat{r}_{t+1}$) is zero. In the mean–variance utility setting, the optimal equity portfolio weight is inversely proportional to the investor's absolute risk aversion (γ) and the risk-free rate (r_{t+1}^f). Similarly, in the power utility setting, the optimal weight (Equation (16)) reduces to $\omega_t^* = 1/(2\gamma)$, where the allocation to equities is entirely driven by the investor's relative risk aversion.

We compare the asset allocation effectiveness of different forecasts using the Sharpe ratio and the certainty equivalent return (CER). The CER is the risk-free rate of return that an investor considers equal (in terms of expected utility) to a different expected return which is higher but also riskier. The CER varies across investors because of their different risk tolerance. Thus, if $\text{CER}_A > \text{CER}_B$, where A and B denote two competing forecasts used as alternative inputs to solve Equation (14) for asset allocation, forecast A entails higher expected utility for the investor than forecast B. The CER of a mean–variance utility investor is given by the average realised utility over the OOS period

$$\text{CER} = J^{-1} \sum_{j=0}^{J-1} U(W_{t+j+1}) = \hat{\mu}_p - \frac{\gamma \hat{\sigma}_p^2}{2}, \tag{17}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the realised mean and variance of the portfolio excess returns over the OOS period, and γ is the investor's absolute risk aversion. The CER of a power utility investor is given by

$$\text{CER} = \left[\frac{(1-\gamma)}{J} \sum_{j=0}^{J-1} \frac{W_{t+j+1}^{1-\gamma}}{1-\gamma} \right]^{1/(1-\gamma)} - 1, \tag{18}$$

where γ is the investor's relative risk aversion, and W_{t+j+1} is as defined in Equation (13).

Our forecast evaluation metric is the CER gain, Δ, defined as the difference between the CER of a investor who employs a model-based forecasting approach and the CER of

a investor that assumes no predictability and, accordingly, simply relies on the historical average. The test of $H_0 : \Delta \leq 0$ against the alternative $H_a : \Delta > 0$ is done using the test in [54], with the critical values estimated by their proposed bootstrap method. We report the annualised Δ in percentage, obtained by multiplying the monthly CER gain by 1200; $\Delta > 0$ can be interpreted as the annualised fee that an investor would be willing to pay in order to have access to the model-based forecast.

5. Empirical Results

Our empirical analysis is based on monthly data from December 1950 to December 2017 for the S&P 500 index and the 12 macroeconomic variables and 14 technical indicators described in Section 2. The equity risk premium, r_t is defined as the return on a broad stock market index, including dividends, from month $t-1$ to month t, in excess of the risk-free interest rate (i.e., 3-month U.S. Treasury bill rate) as at month t. The data source is Amit Goyal's webpage (See https://sites.google.com/view/agoyal145 (accessed on 13 December 2024) for access to various datasets, including more recent ones).

The analysis is based on 1-step-ahead OOS forecasts constructed recursively using expanding windows. The first estimation window spans the 180-month period from December 1950 to December 1965; the second window spans 181 months and so forth, with a total of 224 forecast. Thus, the first forecast $\hat{r}_{181}$ is for January 1966 and the last forecast, $\hat{r}_{404}$ is for December 2017. (In the Section 4 notation $T = 180$ and $J = 224$). Section Abbreviations compiles all the model and forecast-accuracy abbreviations employed in the paper.

5.1. Business-Cycle Dating with Technical Indicators

We begin by showing empirically that the agreement technical indicators constructed from equity market prices and/or volume data serve as real-time business cycle indicators. Table 1 reports the *hits*, or frequency, with which each agreement technical indicator, as defined in Section 2, anticipates an expansion for month $t+1$ and the month is classified as such by the NBER-dating; likewise, for recession months. We also report the frequency of transitions (from expansion to recession or vice versa) according to each technical indicator to compare it with the actual NBER cycle transition, which is 2.61% over the sample months. Overall, the recession (expansion) hits range between 60.8% and 86.3% (53.7% and 90.9%) across technical indicators, and the frequency of transitions between 5.7% and 12.5%. There is no agreement indicator which is the best one according to hits and state transition. Bearing this in mind, the agreement indicator is selected every time we do a forecasting. As an illustration, we plot the $A_{10,t}$ indicator over the entire sample period together with shaded areas that indicate NBER-dated recessions. Figure 1 illustrates that the agreement among technical indicators effectively signals the economic state for the month ahead.

Table 1. Business-cycle dating with technical indicators. The table reports the hits or frequency with which agreement technical indicator computed with information up to month t signals a state (expansion or recession) for month $t+1$, which coincides with the NBER-dating. The sample period is January 1951 to December 2017. It also reports the percentage of transitions from one state to another over the sample months.

	Agreement $A_{\tau,t}$ Indicators Hits (%)			**Transitions**
	All	**Exp**	**Rec**	**(%)**
A_1	55.5	90.9	60.8	12.5
A_2	63.4	87.6	67.0	11.7

Table 1. *Cont.*

	Agreement $A_{\tau,t}$ Indicators Hits (%)			Transitions
	All	**Exp**	**Rec**	**(%)**
A_3	68.2	86.0	70.9	11.0
A_4	71.2	85.1	73.3	10.7
A_5	73.9	83.5	75.4	9.7
A_6	76.4	80.2	77.0	10.0
A_7	80.5	76.9	80.0	8.5
A_8	83.3	76.0	82.2	7.5
A_9	85.4	72.7	83.5	6.2
A_{10}	85.9	69.4	83.5	6.0
A_{11}	87.7	67.8	84.7	6.5
A_{12}	88.4	64.5	84.8	5.7
A_{13}	89.8	61.2	85.4	6.5
A_{14}	92.1	53.7	86.3	7.0

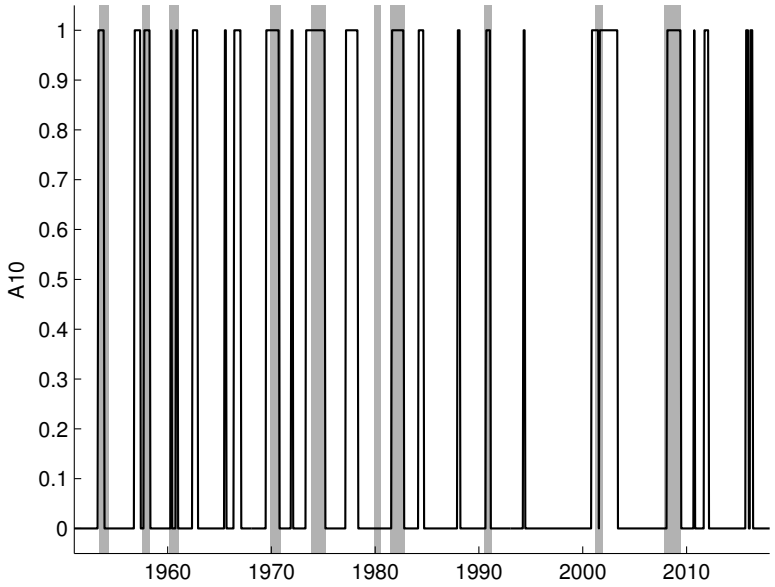


Figure 1. Plot of the recession months as signalled by the agreement (of technical indicators) variable A_{10}. Shaded areas indicate recession months according to NBER business cycle dating. The sample period is from December 1950 to December 2017.

5.2. Results from Statistical Evaluation of Combined Forecasts

Tables 2 and 3 report the results when the forecast covers all the period (January 1966 to December 2017) and from January 1976, respectively. The two periods are similar to the periods considered by [8,13] (the difference is that they use quarterly data and the prediction starts in 1965). The second period is motivated by the results of [8], which suggest that the OOS predictive ability of several macroeconomic variables for the equity premium deteriorates sharply after the dramatic oil price shocks that occurred during the 1973–1975 period. Panel A focuses on LIN models. Panels B and C report the results on TSR models with the agreement technical indicators $A_{p,t}$ as observable. Panels D and E report the performance of the MS models. We also report, in Panel F, linear models that employ as predictors the first three principal components (PC) of either the set of 12 macroeconomic variables, the 14 technical indicators, or all 26 of them. Panels G and H report the results using gradient boosting and random forest, respectively.

Table 2. Statistical and economic evaluation of equity premium forecasts. The table reports statistical (**left** panel) and economic (**right** panel) measures to assess the accuracy of out-of-sample EW combined forecasts over the OOS period from January 1966 to December 2017. The underlying forecasts are from the LIN models (Panel A), two-state threshold regression (TSR) model with the same intercept and different intercepts (Panels B and C, respectively), Markov-switching (MS) model with the same intercept and different intercepts (Panel D and E, respectively), principal component analysis (PCA, Panel F), gradient boosting (Panel G), and random forest (Panel H). The R^2_{OOS} measures the reduction in MSE attained by the model-based forecast at hand relative to the HA benchmark (2–4 columns) or relative to the EW-LIN ($k = 1$) combination of [13] as benchmark (column labelled EW (k = 1)). The hypothesis H_0: $R^2_{OOS} \leq 0$ against $R^2_{OOS} > 0$ is tested using the [51] MSE-adjusted statistic. The table also summarises the asset-allocation value of each predictor for a mean–variance (MV) utility investor with absolute risk aversion parameter of 5, and a CRRA preferences investor with relative risk aversion parameter of 5; $\Delta(\%)$ is the annualised certainty equivalent return (CER) gain versus the HA benchmark. The hypothesis H_0: $\Delta(\%) \leq 0$ against $\Delta(\%) > 0$ is tested using the [54] test. In all tests: *, **, and *** indicate rejection at the 10%, 5%, and 1% significance levels. Bold denotes the best two models and italics denotes the worst two models according to each criterion.

Parameters	R^2_{OOS}				**MV**		**CRRA**	
	Overall	**EXP**	**REC**	**EW ($k = 1$)**	**$\Delta(\%)$**	**Sharpe r.**	**$\Delta(\%)$**	**Sharpe r.**
Panel A: EW-LIN model								
$k = 1$	0.81 ***	0.26	2.03	0.00	1.57 ***	0.111	1.78 ***	0.098
$k = 2$	1.56 ***	**0.95**	2.90	0.75 **	2.50 ***	0.134	2.45 ***	0.115
$k = 3$	**1.72** ***	0.88	3.59	**0.91** **	2.64 ***	0.138	2.62 ***	0.119
Panel B: EW-TSR model—same intercept								
$k = 1$	1.46 **	**0.88**	2.72	0.65 *	2.71 ***	0.137	2.79 ***	0.123
$k = 2$	1.67 **	0.74	3.73	0.86 *	3.25 ***	0.150	3.27 ***	0.134
$k = 3$	**2.00** **	0.53	**5.25**	**1.19** *	**3.32** ***	**0.152**	**3.38** ***	0.135
Panel C: EW-TSR model—different intercept								
$k = 1$	1.23 **	0.27	3.36	0.42	2.87 ***	0.133	2.73 ***	0.125
$k = 2$	1.22 **	0.33	3.18	0.41	3.22 ***	0.144	3.28 **	0.136
$k = 3$	1.01 **	0.03	3.16	0.20	3.20 **	0.146	3.29 **	0.136
Panel D: EW-MS model—same intercept								
$k = 1$	−1.55	−1.31	0.12	−1.36	1.76	0.119	1.88	0.117
$k = 2$	0.50 *	−1.69	3.12	−1.31	2.43	0.128	2.68	0.125
$k = 3$	0.69 **	−1.86	4.09	−1.12	2.62	0.135	3.23	**0.142**
Panel E: EW-MS model—different intercept								
$k = 1$	−1.39	−1.84	−1.42	−1.39	*−1.77*	0.109	*−1.89*	0.099
$k = 2$	0.21	−1.75	2.14	−1.61	2.35	0.123	2.54	0.126
$k = 3$	0.52 *	−1.01	4.10	−1.29	2.33	0.124	2.98	0.132
Panel F: PCA								
ECON	−1.24 **	−1.61	4.98	−1.05 *	1.38 *	0.122	0.63	0.097
TECH	0.53 *	−1.51	2.84	−1.28	2.07 **	0.124	1.96 **	0.107
ALL	1.21 ***	−1.94	**10.36**	0.40 ***	**3.50** ***	**0.160**	**3.40** ***	**0.146**
Panel G: Gradient Boosting n-estimators, learning rate)								
100, 0.01	−1.39	−1.71	−1.68	−1.39	1.47	0.086	1.47	0.094
100, 0.001	−1.25	−1.08	−1.62	−1.25	*0.06*	*0.053*	*−1.05*	*0.060*
200, 0.01	*−1.80*	*−1.60*	*−1.03*	*−1.80*	1.84 *	0.103	1.81	0.105
200, 0.001	−1.68	−1.35	−1.39	−1.68	0.26	*0.057*	0.12	0.095
400, 0.01	*−11.35*	*−13.87*	*−1.79*	*−11.35*	1.49	0.096	1.55	0.100
400, 0.001	−1.86	−1.31	−1.08	−1.86	0.32	0.057	0.13	*0.065*
Panel H: Random Forest (n-estimators, maximum factors)								
100, 2	0.95 **	0.22	2.56	0.14	2.59 **	0.107	2.38 *	0.107
100, 4	−1.24	−1.54	0.42	−1.24	2.69 *	0.118	2.76 *	0.122
200, 2	0.22	−1.26	1.27	0.22	2.98 **	0.125	2.72 **	0.128
200, 4	−1.08	−1.70	1.27	−1.08	2.20 *	0.105	2.23 *	0.109
400, 2	−1.33	−1.64	0.36	−1.33	1.70	0.098	1.93	0.101
400, 4	−1.23	−1.50	0.34	−1.23	2.02	0.111	2.14	0.114

Table 3. Statistical and economic evaluation of post-1976 equity premium forecasts. The table reports statistical (**left** panel) and economic (**right** panel) measures to assess the accuracy of out-of-sample EW combined forecasts over the OOS period from January 1976 to December 2017. The underlying forecasts are from the LIN models (Panel A), two-state threshold regression (TSR) model with the same intercept and different intercepts (Panels B and C, respectively), Markov-switching (MS) model with the same intercept and different intercepts (Panels D and E, respectively), principal component analysis (PCA, Panel F), gradient boosting (Panel G), and random forest (Panel H). The R^2_{OOS} measures the reduction in MSE attained by the model-based forecast at hand relative to the HA benchmark (2–4 columns) or relative to the EW-LIN ($k = 1$) combination of [13] as benchmark (column labelled EW ($k = 1$)). The hypothesis H_0: $R^2_{OOS} \leq 0$ against $R^2_{OOS} > 0$ is tested using the [51] MSE-adjusted statistic. The table also summarises the asset-allocation value of each predictor for a mean–variance (MV) utility investor with absolute risk aversion parameter of 5, and a CRRA preferences investor with relative risk aversion parameter of 5; $\Delta(\%)$ is the annualised certainty equivalent return (CER) gain versus the HA benchmark. The hypothesis H_0: $\Delta(\%) \leq 0$ against $\Delta(\%) > 0$ is tested using the [54] test. In all tests: *, **, and *** indicate rejection at the 10%, 5%, and 1% significance levels. Bold denotes the best two models and italics denotes the worst two models according to each criterion.

Parameters	R^2_{OOS}				**MV**		**CRRA**	
	Overall	**EXP**	**REC**	**EW ($k = 1$)**	**$\Delta(\%)$**	**Shape r.**	**$\Delta(\%)$**	**Sharpe r.**
Panel A: EW-LIN model								
$k = 1$	0.41 *	**−1.02**	1.74	0.00	0.97	0.149	1.15 *	0.136
k = 2	0.31	−1.03	1.33	−1.10	0.15	0.130	0.06	0.111
k = 3	0.19	−1.27	1.57	−1.23	0.03	0.128	−1.03	0.109
Panel B: EW-TSR model—same intercept								
$k = 1$	0.69 *	**0.12**	2.41	0.27	1.54	0.164	1.67 *	0.149
k = 2	0.66 *	−1.42	3.93	0.25	**2.06** **	**0.178**	2.13 **	0.160
k = 3	0.38 *	−1.05	**4.73**	−1.03	1.84 *	**0.172**	2.00 **	0.156
Panel C: EW-TSR model—different intercept								
$k = 1$	**0.88** *	−1.19	4.11	**0.46**	**2.09** **	0.172	**2.41** ***	**0.174**
k = 2	0.61 *	−1.62	4.33	0.19	2.03 **	0.171	**2.43** ***	**0.174**
k = 3	0.24 *	−1.16	4.47	−1.18	1.85 *	0.164	2.11 **	0.166
Panel D: EW-MS model—same intercept								
$k = 1$	−1.07	−1.06	0.82	−1.48	0.94	0.120	0.90	0.120
k = 2	−1.49	−1.82	3.58	−1.90	1.42	0.141	1.67	0.147
k = 3	−1.31	−1.75	1.43	−1.31	1.55	0.145	1.86	0.150
Panel E: EW-MS model—different intercept								
$k = 1$	−1.67	−1.62	−1.78	−1.09	0.95	0.142	0.83	0.142
k = 2	−1.78	−1.93	3.57	−1.19	1.39	0.141	1.60	0.146
k = 3	−1.49	−1.80	1.38	−1.21	1.43	0.146	1.78	0.148
Panel F: PCA								
ECON	−1.59	−1.34	−1.32	−1.01	*−1.52*	0.095	*−1.91*	*0.097*
TECH	0.42	−1.43	3.01	0.01	1.74 *	0.149	1.56	0.150
ALL	−1.72 *	−1.95	**9.06**	−1.13	1.65 *	0.149	1.18	0.146
Panel G: Gradient Boosting (n-estimators, learning rate)								
100, 0.01	−1.51	−1.46	−1.64	−1.93	−1.18	0.100	−1.00	0.099
100, 0.001	−1.27	−1.20	−1.49	−1.69	−1.23	0.095	−1.22	0.108
200, 0.01	*−1.69*	*−10.79*	−1.31	*−1.10*	0.45	0.121	−1.29	0.103
200, 0.001	−1.74	−1.61	−1.14	−1.16	−1.37	*0.092*	0.16	0.113
400, 0.01	*−12.94*	*−15.89*	*−1.00*	*−13.36*	0.01	0.111	0.46	0.125
400, 0.001	−1.30	−1.05	*−1.06*	−1.71	*−1.99*	*0.078*	−1.48	0.100
Panel H: Random Forest (n-estimators, maximum factors)								
100, 2	0.36 *	−1.50	2.96	−1.05	1.43 *	0.140	0.10	0.116
100, 4	**0.71** *	−1.34	3.90	**0.30**	1.80 *	0.151	*−1.21*	*0.086*
200, 2	−1.25	−1.05	2.17	−1.67	1.23	0.135	1.25	0.138
200, 4	−1.49	−1.29	1.92	−1.91	1.08	0.131	0.97	0.131
400, 2	−1.82	−1.63	1.62	−1.24	0.93	0.127	0.87	0.129
400, 4	−1.49	−1.40	2.25	−1.91	1.14	0.132	1.17	0.136

Focusing on the statistical criteria, we now compare the EW combined equity premium forecasts from complete-subset $k = \{1, 2, 3\}$ models within the LIN, TSR, and MS families, and also with the linear model with principal components as explanatory variables and the decision tree methods discussed in Section 3; R^2_{OOS}, the reduction in mean squared error attained by the model-based forecast at hand versus the traditional HA benchmark, Equation (11), is used as statistical performance criterion. It is computed over all the OOS

months and separately over NBER-dated business cycle expansion and recession months. The column labelled EW (k = 1) reports a variant of the R^2_{OOS} statistic using instead as benchmark the EW-LIN ($k = 1$) forecast combination of [13].

Comparing the R^2_{OOS} values for expansion and recession months suggests that the additional value of model-based approaches versus the HA benchmark is more noticeable in recessions than in expansions. During both periods, the gradient boosting method had the worst performance with negative values of R^2_{OOS}. In general, the performance of the MS model, linear model with PC as regressors, and the random forest method are also not good, especially during the expansion periods.

Regarding the linear models, the results not only serve to confirm those in [13] from EW-LIN ($k = 1$) combined forecasts, but also reveal that this is a pervasive finding from the perspective of more general forecasting approaches, such as EW-LIN ($k \in \{2, 3\}$) models. Moreover, relative to the mostly negative R^2_{OOS} values obtained for the individual simple LIN models (results not presented because they are similar to the results presented in [13]), the positive and relatively large R^2_{OOS} attained in Table 2 (Panel A, EW-LIN) reaffirm the usefulness of the forecast combination approach. The forecast performance measures shown in Panel A also reveal in the novel context of the equity premium that the combination of complete-subset 2-predictor EW-LIN ($k = 2$) models or 3-predictor EW-LIN ($k = 3$) models notably improves upon the combination of simple LIN models. In fact, the statistics shown in the fourth column labelled EW ($k = 1$) suggest that the EW-LIN ($k = 2$) and EW-LIN ($k = 3$) predictions significantly reduce the MSE of the EW-LIN ($k = 1$) prediction.

Moving beyond the LIN models, we also appraise the two-state EW-TSR and EW-MS models. The comparison of R^2_{OOS} for a given k across families of models (Panels A to D) suggests that the regime-switching EW-TSR family based on an observable state variable attains superior forecast accuracy gains not only compared to the EW-LIN models but also compared to the EW-MS models based on a latent state variable that are less parsimonious (involving a transition probability matrix) and computationally more demanding. In general, when comparing EW-TSR models with the same or different intercept, none of them clearly dominated the others. We also considered linear models that employ as predictors the first three principal components (PC) of either the set of 12 macroeconomic variables, the 14 technical indicators, or all 26 of them. This type of forecast, known as diffusion indices in the literature, have been shown not to outperform the combined forecasts; see Rapach and Zhou [15] and Neely et al. [9]. The results presented in Panel F shows a good performance in the full sample period. However, the performance is relatively poor in the expansion period, as borne out by the R^2_{OOS}. Moreover, performance during the expansion period generally improves when relying solely on technical indicators.

The main lesson from these results is that constructing equity premium predictions from EW threshold regression models that extend the conventional linear regressions by introducing a predictability state via a technical indicator is very effective.

Figure 2 plots the two components of the equity premium forecast MSE over the entire OOS period, namely, the squared forecast bias and the forecast variance (Theil [55]). The forecast MSE = $\mathrm{E}[(r_{t+j+1} - \hat{r}_{t+j+1})^2]$ can be estimated as $[\sum_{j=0}^{J-1}(\hat{r}_{t+j+1} - r_{t+j+1})/J]^2$, which can be decomposed as the squared forecast bias $[(\sum_{j=0}^{J-1}\hat{r}_{t+j+1} - r_{t+j+1})/J]^2$ and the forecast variance $\sum_{j=0}^{J-1}(\hat{r}_{t+j+1} - \bar{\hat{r}})^2/J$, where $\bar{\hat{r}}$ is the mean of $\hat{r}_{t+j+1}, j = 0, \ldots, (J-1)$.) We plot the results for the simple LIN models, the EW combinations thereof, as well as the EW combinations of the two-state TSR and MS models. To keep the graph readable, we just display the results for the forecast combination of models with one macroeconomic predictor ($k = 1$). The results suggest that the combined forecasts (from either LIN or TSR models) have lower variance than the forecasts from all of the individual LIN models.

Finally, the graph also reveals that the TSR models that allow for regimes of equity premium predictability have lower forecast variance than the LIN models.

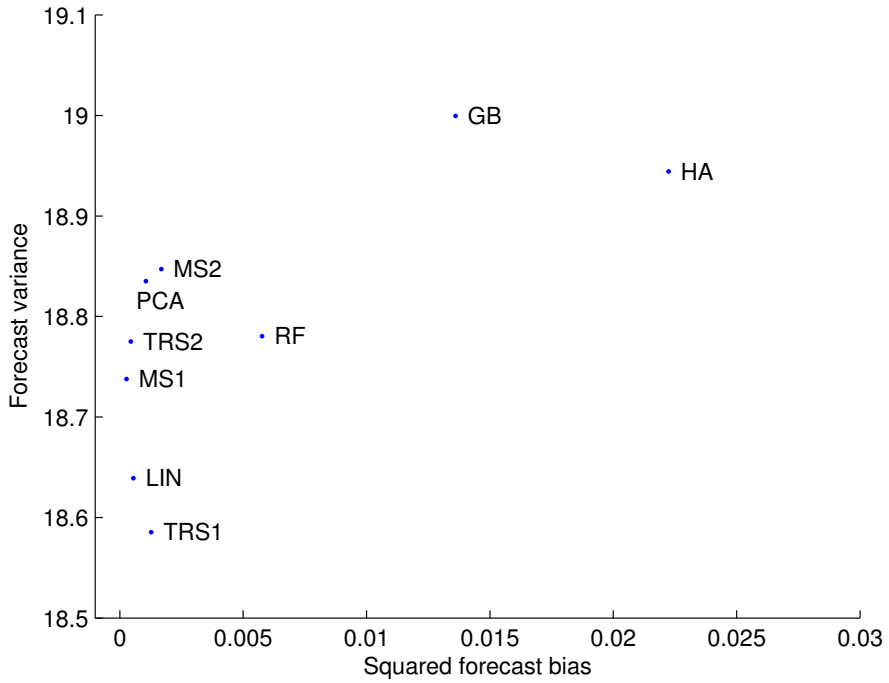


Figure 2. The figure provides a scatterplot of the out-of-sample forecast variance against the squared forecast bias for the combination of linear models (LIN), two-state threshold regressions with the same (TRS1) and different (TRS2) intercepts, two-state Markov-switching with the same (MS1) and different (MS2) intercepts, principal-component analysis (PCA), random-forest (RF), gradient boosting (GB), and historical average (HA). For each type of model, the forecasts being combined are those arising from the complete-subset specifications with one predictor ($k = 1$). The OOS period is from January 1966 to December 2017.

To appraise the OOS forecast performance dynamically over time, we plot in Figure 3, left column, the CSE discussed in Section 3 for the HA benchmark minus the CSE for combined forecasts. We present the results for EW-LIN and EW-TSR models. The EW-TSR model is with the same intercept. We observe that the predictive ability of EW-LIN combined forecasts begins to deteriorate in the late 1990s. On the other hand, the graphs corresponding to the EW-TSR forecasts are predominantly upward sloped. Unreported graphs for the EW-MS model show that their performance begins to deteriorate as early as 1975, while for diffusion indices (i.e., linear forecasts based on the first three principal components of all macroeconomic variables and/or technical indicators) show a dropoff from the mid to late 1970s, a sharp increase between 1980 and 1985, followed by a significant drop until the late 1990s, and another rise thereafter. This evidence suggests that diffusion indices behave more erratically than the EW-LIN forecasts, in line with the findings in [15]. Overall, the evidence from this dynamic forecast evaluation reinforces our early findings from the static evaluation. Specifically, it confirms that the forecast combination from TSR models that parsimoniously allow for two predictive states (expansion versus recession) through a technical indicator consistently outperform the combination of forecasts from simple LIN models. Finally, Figure 3, right column, presents the CSE of the EW-LIN ($k = 1$) combination of [13] minus the CSE of the EW-LIN ($k = 2$) or EW-LIN ($k = 3$) combinations in the first graph, and the EW-TSR combinations in the second graph. The results are qualitatively similar when using the HA as the forecast benchmark.

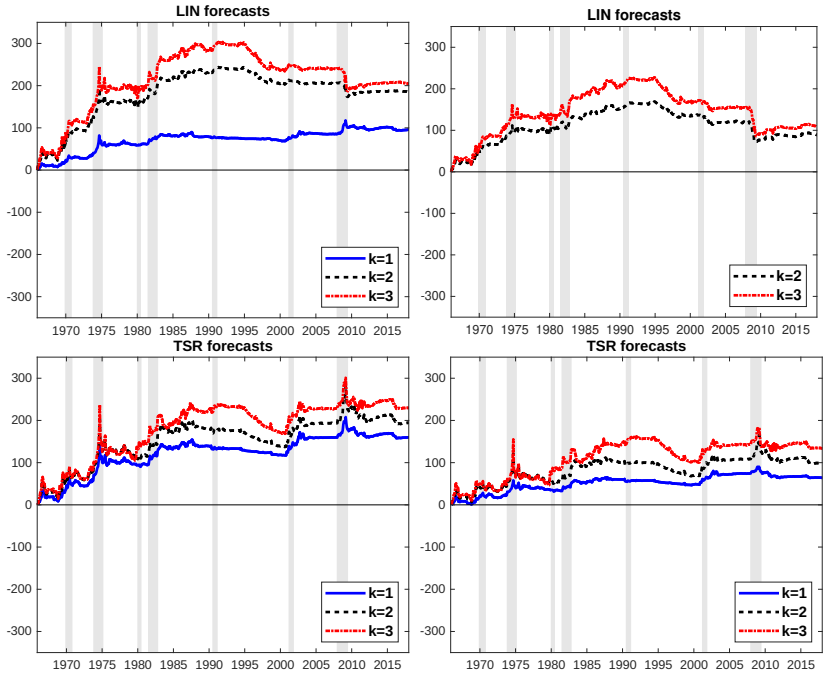


Figure 3. Plots of ΔCSE_t defined as the cumulative squared forecast errors of the historical average (HA) benchmark model minus the cumulative squared forecast errors from combinations of forecasting methods (on the **left**) and the cumulative squared forecast errors of the combining EW-LIN ($k = 1$) combination of the [13] benchmark model minus the cumulative squared forecast errors from the combination of forecasting methods (on the **right**) over the out-of-sample period from January 1966 to December 2017. We consider EW-LIN model and EW-TSR models with the same intercept as predictive models. The candidate predictors are the set of 12 macroeconomic variables considered either in single-, two-, or three-variable regressions ($k = 1, 2, 3$). Shaded areas indicate recession months according to NBER business cycle dating.

5.3. Results from Economic Evaluation of Combined Forecasts

It is well known that the forecast rankings stemming from utility-based criteria and statistical criteria do not necessarily come hand in hand (see, e.g., [15,56], inter alia). Thus, for completeness, we present in Tables 2 and 3, for the whole and crisis periods, respectively, the results from the asset-allocation exercise outlined in Section 4.2 for a representative mean–variance utility investor with absolute risk aversion parameter $\delta = 5$ and a CRRA preferences investor with relative risk aversion parameter $\delta = 5$.

Concurrent with the evidence from the statistical forecast evaluation, the portfolio analysis suggests that the gradient boosting is not very effective, whereas the forecast combination based on TSR models is the most fruitful. However, the performance of the combination of the MS models and linear models are now similar. To illustrate, for the full period, the TSR models for $k = 3$ provide sizeable CER gains of 3.32 and 3.20 in Panels B and C, respectively, versus 2.64 (LIN models in Panel A) and 2.62 and 2.33 for MS models in Panel D and E, respectively. A similar result occurs for a CRRA utility investor. Thus, for both the statistical and economic criteria, the results indicate that the agreement technical indicators $A_{p,t}$ are very effective state variables for equity premium prediction. Regarding the linear models, in Panel A, an analysis reveals that the forecast combination of LIN ($k = 1$) models is improved upon by the forecast combination of complete-subset LIN ($k = 2$) and LIN ($k = 3$) models. For instance, for the full period, with a $\Delta(\%)$ of 2.64 (mean–variance investor) and 2.62 (CRRA investor), the EW-LIN ($k = 3$) combined forecasts deliver a superior CER gain versus the HA benchmark compared to the EW-LIN ($k = 1$) combined forecasts with smaller $\Delta(\%)$ values of 1.57 and 1.78 for mean–variance

and CRRA investors, respectively. The improvement using complete-subset $k = \{2,3\}$ combinations compared to complete-subset $k = 1$ combinations also happens with TSR and MS models (see panels B to E). For the crisis period, the conclusions are similar.

5.4. Robustness Checks

We carry out several robustness checks to assess if the above results are challenged, first when we consider only most recent period (similar to [13], the forecast period starts on January 2000). This third period, 2000–2017, gives relatively more weight to the late 2000s global financial crisis. The statistical and economic results in Table 4 confirms most of the previous results. In particular, EW-TSR models outperform the combination of LIN models for any $k = \{1,2,3\}$, except in terms of R^2_{OOS} for the case without intercept and $k = 3$. The PCA method has also a good performance when the components are evaluated considering either the technical indicators or jointly the macroeconomic and technical indicators. However, when considering only the macroeconomic variables, the performance is worse than the linear models.

Table 4. Statistical and economic evaluation of post-2000 equity premium forecasts. The table reports statistical (**left** panel) and economic (**right** panel) measures to assess the accuracy of out-of-sample EW combined forecasts over the OOS period from January 2000 to December 2017. The underlying forecasts are from the LIN models (Panel A), two-state threshold regression (TSR) model with the same intercept and different intercepts (Panels B and C, respectively), Markov-switching (MS) model with the same intercept and different intercepts (Panels D and E, respectively), principal component analysis (PCA, Panel F), gradient boosting (Panel G), and random forest (Panel H). The R^2_{OOS} measures the reduction in MSE attained by the model-based forecast at hand relative to the HA benchmark (2–4 columns) or relative to the EW-LIN ($k = 1$) combination of [13] as benchmark (column labelled EW (k = 1)). The hypothesis H_0: $R^2_{OOS} \leq 0$ against $R^2_{OOS} > 0$ is tested using the [51] MSE-adjusted statistic. The table also summarises the asset-allocation value of each predictor for a mean–variance (MV) utility investor with absolute risk aversion parameter of 5 and a CRRA preferences investor with relative risk aversion parameter of 5; $\Delta(\%)$ is the annualised certainty equivalent return (CER) gain versus the HA benchmark. The hypothesis H_0: $\Delta(\%) \leq 0$ against $\Delta(\%) > 0$ is tested using the [54] test. In all tests: *, **, and *** indicate rejection at the 10%, 5%, and 1% significance levels. Bold denotes the best two models and italics denotes the worst two models according to each criterion.

Parameters	R^2_{OOS}				**MV**		**CRRA**	
	Overall	**EXP**	**REC**	**EW ($k = 1$)**	**Δ(%)**	**Shape r.**	**Δ(%)**	**Sharpe r.**
Panel A: EW-LIN model								
$k = 1$	0.66	0.25	1.42	0	1.87	0.081	2.48 *	0.093
$k = 2$	−1.49	0.48	−1.30	−1.15	*−1.80*	*0.024*	*−1.11*	*0.038*
$k = 3$	−1.82	0.78	−1.78	−1.48	−1.58	0.031	−1.04	0.043
Panel B: EW-TSR model—same intercept								
k = 1	1.09 *	1.08	1.11	0.43	3.04 ***	0.105	3.67 ***	0.110
k = 2	1.15 *	0.65	2.09	0.49	4.01 ***	0.152	**4.54** ***	0.148
k = 3	0.75	0.00	2.14	0.09	4.15 ***	**0.155**	4.30 ***	0.144
Panel C: EW-TSR model—different intercept								
k = 1	1.54 **	0.07	4.25	0.88	2.99 **	0.112	4.20 ***	0.121
k = 2	1.02 *	−1.16	3.20	0.36	4.02 ***	0.147	4.13 ***	0.139
k = 3	0.34	−1.64	2.14	−1.32	**4.22** ***	0.150	3.99 **	**0.153**
Panel D: EW-MS model—same intercept								
k = 1	−1.01	*−1.91*	0.02	−1.67	1.23	0.120	1.45	0.124
k = 2	0.83	−1.47	1.98	0.17	1.94	0.141	2.26	0.150
k = 3	1.36 *	0.33	2.39	0.70	3.85	0.145	3.46	0.151
Panel E: EW-MS model—different intercept								
k = 1	*−1.51*	*−1.12*	−1.11	*−1.17*	0.58	0.129	1.69	0.138
k = 2	0.94	−1.12	2.00	0.28	1.93	0.134	3.11	0.147
k = 3	1.22	0.45	2.19	0.56	2.24	0.148	3.35	0.148

Table 4. *Cont.*

Parameters	R^2_{OOS}				MV		CRRA	
	Overall	EXP	REC	EW ($k = 1$)	Δ(%)	Shape r.	Δ(%)	Sharpe r.
Panel F: PCA								
ECON	*−1.05*	−1.96	*−1.90*	*−1.71*	*−1.73*	*0.028*	*−1.48*	*0.008*
TECH	**1.55** *	0.01	**4.38**	**0.89**	4.09 ***	**0.157**	4.35 ***	0.152
ALL	**1.72** **	−1.55	**5.90**	**1.06**	3.98 ***	0.149	**4.75** ***	**0.162**
Panel G: Gradient Boosting (n-estimators, learning rate)								
100, 0.01	−1.29	**2.11**	*−1.73*	−1.95	1.19	0.129	1.99	0.128
100, 0.001	−1.12	0.12	−1.55	−1.78	2.30	0.114	2.02	0.115
200, 0.01	0.65	**3.06**	−1.80	−1.01	2.22	0.122	2.73	0.101
200, 0.001	−1.18	0.40	−1.27	−1.84	0.99	0.103	2.00	0.092
400, 0.01	−1.46	0.00	−1.14	−1.12	1.92	0.126	1.55	0.125
400, 0.001	−1.72	0.94	−1.78	−1.38	0.12	0.081	2.98 *	0.095
Panel H: Random Forest (n-estimators, maximum factors)								
100, 2	1.38 *	0.50	3.00	0.72	3.97 **	0.129	3.76 **	0.129
100, 4	1.52 *	0.67	3.08	0.86	**4.25** ***	0.139	4.09 **	0.139
200, 2	0.42	0.66	−1.02	−1.24	2.91 *	0.094	2.57 *	0.097
200, 4	0.63	0.43	0.98	−1.03	2.78 *	0.090	2.29	0.089
400, 2	0.49	0.78	−1.03	−1.17	3.18 **	0.103	2.80 *	0.104
400, 4	0.82	0.90	0.68	0.16	3.03 *	0.098	2.80 *	0.104

Second, in the spirit of [57], we also evaluate the asset allocation exercise by considering a CRRA parameter equal to 3 and 10. The results, presented in Table 5, confirm the better performance of EW-TSR models.

Table 5. Asset allocation exercise for $\gamma = 3$ and $\gamma = 10$: The table reports economic (**right** panel) measures to assess the accuracy of out-of-sample EW combined forecasts over the OOS period from January 1966 to December 2017. The underlying forecasts are from the LIN models (Panel A), two-state threshold regression (TSR) model with the same intercept and different intercepts (Panels B and C, respectively), the Markov-switching (MS) model with the same intercept and different intercepts (Panels D and E, respectively), principal component analysis (PCA, Panel F), gradient boosting (Panel G), and random forest (Panel H). The investor who allocates his wealth between stocks and risk-free bills at the end of each month is assumed to have a mean–variance or constant relative risk aversion (CRRA) preferences and a relative risk aversion parameters $\gamma = 3$ (**left** panel) and $\gamma = 10$ (**right** panel); $\Delta(\%)$ is the annualised certainty equivalent return (CER) gain versus the HA benchmark. The hypothesis H_0: $\Delta(\%) \leq 0$ against $\Delta(\%) > 0$ is tested using the [54] test: *, **, and *** indicate rejection at the 10%, 5%, and 1% significance levels. Bold denotes the best two models and italics denotes the worst two models according to each criterion.

Parameters	MV		CRRA		MV		CRRA	
	Δ(%)	Shape r.	Δ(%)	Sharpe r.	Δ(%)	Shape r.	Δ(%)	Sharpe r.
Panel A: EW-LIN model								
$k = 1$	1.38 *	0.114	1.53	0.091	1.08	0.110	1.13	0.106
$k = 2$	3.11 **	0.150	3.52 **	0.126	1.31	0.122	1.30	0.116
$k = 3$	3.23 ***	0.152	**3.72** ***	0.131	1.20	0.120	1.20	0.115
Panel B: EW-TSR model—same intercept								
$k = 1$	2.75 **	0.135	2.83 **	0.124	1.53 *	0.131	1.55 **	0.126
$k = 2$	**3.62** ***	**0.153**	3.70 ***	0.137	**1.84** **	0.147	**1.93** **	0.143
$k = 3$	3.35 ***	0.147	3.49 **	0.135	1.66 **	0.142	1.71 **	0.138
Panel C: EW-TSR model—different intercept								
$k = 1$	3.17 **	0.133	3.47 ***	0.133	1.56 *	0.138	1.56 *	0.138
$k = 2$	**3.59** ***	0.143	3.59 ***	**0.138**	**1.84** **	**0.148**	**1.94** **	**0.148**
$k = 3$	3.58 ***	0.146	3.58 ***	0.141	1.80 **	0.146	1.80 **	**0.146**
Panel D: EW-MS model—same intercept								
$k = 1$	1.12	0.110	1.68	0.097	0.62	0.111	1.12	0.103
$k = 2$	1.92	0.131	2.03	0.128	1.25	0.127	1.49	0.118
$k = 3$	2.16	0.128	2.04	0.130	1.65	0.125	1.73	0.123

Table 5. *Cont.*

Parameters	MV		CRRA		MV		CRRA	
	Δ(%)	Shape r.	Δ(%)	Sharpe r.	Δ(%)	Shape r.	Δ(%)	Sharpe r.
Panel E: EW-MS model—different intercept								
$k = 1$	0.99	0.095	*−1.05*	0.081	*−1.67*	0.088	*−1.06*	0.080
$k = 2$	1.68	0.127	2.11	0.125	1.26	0.124	1.78	0.122
$k = 3$	1.84	0.129	1.98	0.129	1.33	0.127	1.55	0.119
Panel F: PCA								
ECON	2.67	0.137	2.62	0.111	*−1.19*	0.100	*−1.42*	0.083
TECH	2.35 *	0.135	2.65 *	0.111	1.18	0.122	1.07	0.113
ALL	3.59 ***	**0.159**	**4.02** ***	**0.143**	1.66 **	**0.152**	1.18	0.145
Panel G: Gradient Boosting (n-estimators, learning rate)								
100, 0.01	2.40	0.106	2.27	0.105	0.20	0.061	−1.25	0.075
100, 0.001	*0.10*	*0.064*	*−1.06*	*0.065*	0.06	0.048	0.03	*0.063*
200, 0.01	2.65 *	0.111	2.30	0.105	0.17	0.079	−1.18	0.088
200, 0.001	*0.32*	*0.068*	0.11	*0.068*	0.10	*0.048*	0.05	0.063
400, 0.01	2.84 *	0.115	2.74 *	0.112	−1.24	0.074	−1.69	0.082
400, 0.001	0.69	0.074	0.37	0.072	0.09	*0.047*	−1.04	*0.062*
Panel H: Random Forest (n-estimators, maximum factors)								
100, 2	3.31 **	0.124	2.99 *	0.117	1.44	0.110	1.47	0.120
100, 4	3.57 **	0.130	3.41 **	0.125	1.40	0.112	1.38	0.121
200, 2	3.32 **	0.124	3.05 *	0.119	1.40	0.109	1.44	0.119
200, 4	2.90	0.116	2.61	0.110	1.23	0.100	1.24	0.110
400, 2	2.98 *	0.117	2.71	0.112	1.31	0.105	1.34	0.115
400, 4	3.40 **	0.126	3.21 *	0.122	1.43	0.110	1.49	0.121

6. Summary and Concluding Remarks

The equity premium forecast combination approaches deployed by [13,15,40] employ both EW and weights based on the past forecasting performance of the individual linear models. We propose extending this EW prediction by combining state-dependent models that interact macroeconomic predictors with a binary technical indicator as an observed state variable to capture the state of the economy. The OOS forecast accuracy of the models is compared with that of existing equity premium forecasting strategies that combine with EW simple linear models. We also consider EW from Markov-switching models. The evaluation of equity premium forecasts is based not only on statistical criteria but also on economic criteria such as the asset-allocation problem of an investor. The analysis is based on monthly data from December 1950 to December 2017 for the S&P 500 equity index. The findings suggest that the proposed approach contributes to delivering stable OOS gains. The threshold regression models offer superior forecast accuracy to existing linear (one-state) specifications and also to more heavily parameterised and computationally demanding Markov-switching models where the state variable is latent. In addition, we found an improvement using complete-subset $k = \{2, 3\}$ combinations compared to complete-subset $k = 1$ combinations with LIN, TSR, and MS models (see panels B to E). For the crisis period, the conclusions are similar. Finally, the machine learning methods, random forest, and gradient boosting presented a worse performance than the combined traditional regression models. These findings are not challenged by different forecasting sub-periods, alternative investor's risk aversion levels, and rolling (versus expanding) estimation windows.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math13020257/s1.

Author Contributions: Conceptualization, D.d.A., A.-M.F., and L.K.H.; methodology, D.d.A., A.-M.F., and L.K.H.; software, D.d.A.; validation, D.d.A., A.-M.F., and L.K.H.; formal analysis, D.d.A., A.-M.F., and L.K.H.; investigation, D.d.A., A.-M.F., and L.K.H.; resources, D.d.A., A.-M.F., and L.K.H.; data curation, D.d.A.; writing—original draft preparation, D.d.A., A.-M.F., and L.K.H.; writing—review and editing, D.d.A., A.-M.F., and L.K.H.; supervision, A.-M.F. and L.K.H.; project administration, A.-M.F. and L.K.H.; funding acquisition. D.d.A. and L.K.H. All authors have read and agreed to the published version of the manuscript.

Funding: Daniel acknowledges financial support from the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), grant 0969/13-3. Luiz acknowledges financial support from CAPES, grant 10600/13-2 and São Paulo Research Foundation (FAPESP), grants 2013/00506-1, 2023/01728-0, and 2023/02538.

Data Availability Statement: The original contributions presented in this study are included in the supplementary material. Further inquiries can be directed to the corresponding author.

Acknowledgments: We thank the participants at the 2016 Austrian Statistics Days, Vienna, the 2016 Workshop of Time Series, Wavelets and Functional Data Analysis at the Institute of Mathematics and Statistics, University of São Paulo, and the 2017 10th SoFiE conference, NYU Stern School of Business, for their comments. Daniel and Luiz thank Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS). We also thank two anonymous referees from Mathematics.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript for the models:

A_τ^{TECH}	Agreement technical indicators, function of the 14 technical indicators
EW	Equally weighted combination of all forecasts from complete-subset models
EW-LIN	Equally weighted combination of forecasts from linear models
EW-MS	Equally weighted combination of forecasts from MS models
EW-TR	Equally weighted combination of forecasts from TR models
LIN	Linear regression model
MA(a, b)	Moving-average technical indicator with a and b months
MOM(l)	Momentum technical indicator with l months
MS	Two-state Markov-switching regression model with latent state variable
TSR	Two-state threshold regression model with A_τ^{TECH} as observable state variable
VOL(a, b)	Volume- and price-based technical indicator with a and b months
Forecast evaluation	
CER	Certainty equivalent return or risk-free return that an investor considers equal (in terms of expected utility) to a higher but risky expected return
CRRA	Constant relative risk aversion or power utility investor
CSE	Cumulative square error of HA minus cumulative square error of forecast at hand
HA	Historical average (as benchmark forecast)
MSE	Mean squared (prediction) error
MV	Mean–variance utility investor
OOS	Out-of-sample
$\Delta(\%)HA$	CER of forecast at hand minus CER of historical average benchmark
$\Delta(\%)EW$	CER of forecast at hand minus CER of EW-LIN ($k = 1$) forecast benchmark

References

1. Fama, E.F.; French, K.R. Dividend yields and expected stock returns. *J. Financ. Econ.* **1988**, *22*, 3–25. [CrossRef]
2. Lettau, M.; Ludvigson, S. Consumption, aggregate wealth, and expected stock returns. *J. Financ.* **2001**, *56*, 815–849. [CrossRef]
3. Ang, A.; Bekaert, G. International asset allocation with regime shifts. *Rev. Financ. Stud.* **2002**, *15*, 1137–1187. [CrossRef]
4. Ang, A.; Bekaert, G. Stock return predictability: Is it there? *Rev. Financ. Stud.* **2007**, *20*, 651–707. [CrossRef]
5. Bossaerts, P.; Hillion, P. Implementing statistical criteria to select return forecasting models: What do we learn? *Rev. Financ. Stud.* **1999**, *12*, 405–428. [CrossRef]
6. Goyal, A.; Welch, I. Predicting the equity premium with dividend ratios. *Manag. Sci.* **2003**, *49*, 639–654. [CrossRef]
7. Butler, A.W.; Grullon, G.; Weston, J.P. Can managers forecast aggregate market returns? *J. Financ.* **2005**, *60*, 963–986. [CrossRef]
8. Welch, I.; Goyal, A. A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.* **2008**, *21*, 1455–1508. [CrossRef]
9. Neely, C.J.; Rapach, D.E.; Tu, J.; Zhou, G. Forecasting the equity risk premium: The role of technical indicators. *Manag. Sci.* **2014**, *60*, 1772–1791. [CrossRef]

10. Fernandez-Perez, A.; Fuertes, A.M.; Miffre, J. Commodity Markets, Long-Run Predictability, and Intertemporal Pricing. *Rev. Financ.* **2017**, *21*, 1159–1188. [CrossRef]
11. Batten, J.A.; Kinateder, H.; Wagner, N. Beating the average: Equity premium variations, uncertainty, and liquidity. *Abacus* **2022**, *58*, 567–588. [CrossRef]
12. Campbell, J.Y.; Thompson, S.B. Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financ. Stud.* **2008**, *21*, 1509–1531. [CrossRef]
13. Rapach, D.E.; Strauss, J.K.; Zhou, G. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Rev. Financ. Stud.* **2010**, *23*, 821–862. [CrossRef]
14. Ferreira, M.A.; Santa-Clara, P. Forecasting stock market returns: The sum of the parts is more than the whole. *J. Financ. Econ.* **2011**, *100*, 514–537. [CrossRef]
15. Rapach, D.E.; Zhou, G. Forecasting stock returns. In *Handbook of Economic Forecasting*; Elliott, G., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2013; Volume 2, pp. 328–383.
16. Pettenuzzo, D.; Timmermann, A.; Valkanov, R. Forecasting stock returns under economic constraints. *J. Financ. Econ.* **2014**, *114*, 517–553. [CrossRef]
17. Chicaroli, R.; Valls Pereira, P.L. Predictability of Equity Models. *J. Forecast.* **2015**, *34*, 427–440. [CrossRef]
18. Ciner, C. Predicting the equity market risk premium: A model selection approach. *Econ. Lett.* **2022**, *215*, 110448. [CrossRef]
19. Lima, L.R.; Godeiro, L.L. Equity-premium prediction: Attention is all you need. *J. Appl. Econom.* **2023**, *38*, 105–122. [CrossRef]
20. Lu, F.; Ma, F.; Guo, Q. Less is more? New evidence from stock market volatility predictability. *Int. Rev. Financ. Anal.* **2023**, *89*, 102819. [CrossRef]
21. Goyal, A.; Welch, I.; Zafirov, A. A comprehensive 2022 look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.* **2024**, *37*, 3490–3557. [CrossRef]
22. Merton, R.C. An intertemporal capital asset pricing model. *Econometrica* **1973**, *41*, 867–887. [CrossRef]
23. Cochrane, J.H. Presidential address: Discount rates. *J. Financ.* **2011**, *66*, 1047–1108. [CrossRef]
24. Bali, T.G. The intertemporal relation between expected returns and risk. *J. Financ. Econ.* **2008**, *87*, 101–131. [CrossRef]
25. Turner, C.M.; Startz, R.; Nelson, C.R. A Markov model of heteroskedasticity, risk, and learning in the stock market. *J. Financ. Econ.* **1989**, *25*, 3–22. [CrossRef]
26. Garcia, R.; Perron, P. An analysis of the real interest rate under regime shifts. *Rev. Econ. Stat.* **1996**, *78*, 111–125. [CrossRef]
27. Perez-Quiros, G.; Timmermann, A. Firm size and cyclical variations in stock returns. *J. Financ.* **2000**, *55*, 1229–1262. [CrossRef]
28. Ang, A.; Chen, J. Asymmetric correlations of equity portfolios. *J. Financ. Econ.* **2002**, *63*, 443–494. [CrossRef]
29. Guidolin, M.; Timmermann, A. An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *J. Appl. Econom.* **2006**, *21*, 1–22. [CrossRef]
30. Guidolin, M.; Timmermann, A. Term structure of risk under alternative econometric specifications. *J. Econom.* **2006**, *131*, 285–308. [CrossRef]
31. Timmermann, A. Forecast combinations. In *Handbook of Economic Forecasting*; Granger, J., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, pp. 135–196.
32. Pettenuzzo, D.; Timmermann, A. Predictability of stock returns and asset allocation under structural breaks. *J. Econom.* **2011**, *164*, 60–78. [CrossRef]
33. Tu, J. Is regime switching in stock returns important in portfolio decisions? *Manag. Sci.* **2010**, *56*, 1198–1215. [CrossRef]
34. Henkel, S.J.; Martin, J.S.; Nardari, F. Time-varying short-horizon predictability. *J. Financ. Econ.* **2011**, *99*, 560–580. [CrossRef]
35. Dangl, T.; Halling, M. Predictive regressions with time-varying coefficients. *J. Financ. Econ.* **2012**, *106*, 157–181. [CrossRef]
36. Gargano, A.; Timmermann, A. Predictive dynamics in commodity prices. *Int. J. Forecast.* **2014**, *30*, 825–843. [CrossRef]
37. Jacobsen, B.; Marshall, B.R.; Visaltanachoti, N. Stock Market Predictability and Industrial Metal Returns. Technical Report, 23rd Australasian Finance and Banking Conference 2010 Paper, 2016. Available online: https://ssrn.com/abstract=1660864 (accessed on 12 September 2018). [CrossRef]
38. Guidolin, M.; Timmermann, A. Asset allocation under multivariate regime switching. *J. Econ. Dyn. Control* **2007**, *31*, 3503–3544. [CrossRef]
39. Zhu, X.; Zhu, J. Predicting stock returns: A regime-switching combination approach and economic links. *J. Bank. Financ.* **2013**, *37*, 4120–4133. [CrossRef]
40. Elliott, G.; Gargano, A.; Timmermann, A. Complete subset regressions. *J. Econom.* **2013**, *177*, 357–373. [CrossRef]
41. Fama, E.F.; French, K.R. Business conditions and expected returns on stocks and bonds. *J. Financ. Econ.* **1989**, *25*, 23–49. [CrossRef]
42. Mele, A. Asymmetric stock market volatility and the cyclical behavior of expected returns. *J. Financ. Econ.* **2007**, *86*, 446–478. [CrossRef]
43. Harvey, D.I.; Leybourne, S.J.; Sollis, R.; Taylor, A.R. Real-time detection of regimes of predictability in the US equity premium. *J. Appl. Econom.* **2021**, *36*, 45–70. [CrossRef]

44. Stein, T. Forecasting the equity premium with frequency-decomposed technical indicators. *Int. J. Forecast.* **2024**, *40*, 6–28. [CrossRef]
45. Gu, S.; Kelly, B.; Xiu, D. Empirical asset pricing via machine learning. *Rev. Financ. Stud.* **2020**, *33*, 2223–2273. [CrossRef]
46. Xu, X.; Liu, W.h. Forecasting the equity premium: Can machine learning beat the historical average? *Quant. Financ.* **2024**, *24*, 1445–1461. [CrossRef]
47. Chauvet, M.; Piger, J. A comparison of the real-time performance of business cycle dating methods. *J. Bus. Econ. Stat.* **2008**, *26*, 42–49. [CrossRef]
48. Perlin, M. MS Regress the MATLAB Package for Markov Regime Switching Models (2012). Technical Report, 2014. Available online: http://ssrn.com/abstract=1714016 (accessed on 8 August 2018). [CrossRef]
49. Leippold, M.; Wang, Q.; Zhou, W. Machine learning in the Chinese stock market. *J. Financ. Econ.* **2022**, *145*, 64–82. [CrossRef]
50. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **2001**, *29*, 1189–1232. [CrossRef]
51. Clark, T.E.; West, K.D. Approximately normal tests for equal predictive accuracy in nested models. *J. Econom.* **2007**, *138*, 291–311. [CrossRef]
52. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–265. [CrossRef]
53. West, K.D. Asymptotic inference about predictive ability. *Econometrica* **1996**, *30*, 1067–1084. [CrossRef]
54. McCracken, M.W.; Valente, G. Asymptotic inference for performance fees and the predictability of asset returns. *J. Bus. Econ. Stat.* **2018**, *36*, 426–437. [CrossRef]
55. Theil, H. *Applied Economic Forecasting*; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1971.
56. Andrada-Felix, J.; Fernandez-Rodriguez, F.; Fuertes, A.M. Combining nearest neighbor predictions and model-based predictions of realized variance: Does it Pay? *Int. J. Forecast.* **2016**, *32*, 695–715. [CrossRef]
57. Cenesizoglu, T.; Timmermann, A. Do return prediction models add economic value? *J. Bank. Financ.* **2012**, *36*, 2974–2987. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.