Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Combining Combinations: Improving Equity Premium Predictions through Bagging and Forecast Combination.

Mathijs Willems (608651)

Supervisor:	BP van der Sluis
Second assessor	Dr. A. Tetereva
Date final version:	1st July 2024

Abstract

Accurately forecasting the equity premium out-of-sample has proven difficult over the past decades partly because of the challenges posed by structural instability, the complicated characteristics of the data generating process and numerous possible predictors. Bootstrap aggregating (bagging) and forecast combination can be employed to reduce the adverse effects of some of these obstacles and have been applied separately to produce excess stock returns that are significantly more accurate than the historical average. This paper investigates whether applying these techniques in tandem can improve the accuracy of monthly one-step-ahead equity premium predictions. In addition to combining forecasts over the whole set of predictors, a subset combination approach based on automatic rolling-window selection is evaluated. The results show that this sequential forecasting method outperforms the historical average by a substantial margin for short in-sample estimation windows, as bagging reduces the mean squared prediction error and forecast combination offers a decrease in accuracy risk.

1 Introduction

Predicting excess returns has become a central topic in the finance literature due to the economic value that accurate forecasts can provide to investors. Numerous financial macroeconomic variables have been proposed as predictors for the equity premium, however Welch and Goyal (2008) find that most of these explanatory variables fail to outperform the historical average out-of-sample when used in a simple linear model. Some of the major challenges in producing equity premium forecasts with improved accuracy relative to the historical average are structural instability, the complicated characteristics of the data generating process (DGP) and the numerous possible predictors. Bagging was introduced to improve forecasting accuracy by smoothing instabilities and can be similarly utilised for equity premium forecasting. Even though it was originally intended to be applied to independent and identically distributed (IID) data, Jin, Su and Ullah (2014) proof that teir revised version reduces the mean squared prediction error (MSPE) of time series forecasts. Additionally, they find that forecasts based on macroeconomic variables can outperform the historical average when using bagging. Forecast combination allows forecasters to more accurately model the intricate characteristics of the data generating process, mitigate the instability risk associated with the reliance on a single model and incorporate information from different forecasting models. Rapach, Strauss and Zhou (2010) find that combining produces statistically and economically improved equity premium forecasts compared to the historical average

Although both of the previously described methods have been employed separately to improve excess stock return predictions and Rapach and Strauss (2010) find that in some instances forecasts based on both methods contain different information, there is, to the best of our knowledge, no literature that investigates their joint application to equity premium forecasting. Therefore this paper will look into the question: Can the accuracy of monthly one-step-ahead equity premium forecasts be improved by applying bagging and forecast combination in tandem?

Before employing both techniques sequentially to forecast the equity premium, the performance of bagging in a time series setting is first evaluated using Monte Carlo simulations. Two bootstrap aggregating methods are considered in the form of traditional bagging and the revised method, introduced by (Jin et al., 2014), which makes use of historical forecasts to improve predictions. These techniques are applied to parametric as well as non-parametric forecasting models, three DGPs and numerous parameter settings. Consequently, the simulations provide comprehensive insight into the accuracy of the bagging methods relative to the non-bagged forecasts and each other in various scenarios.

To address the main research question, both bagging and forecast combination are used to produce monthly equity premium forecasts based on macroeconomic variables. Similar to the Monte Carlo simulations, the traditional and revised bagging methods are applied to forecasts produced by parametric and non-parametric models. Following this, both the bagged and non-bagged forecasts are combined across the different predictive variables. These combinations are constructed using the whole set of predictors as well as a subset of variables with the best historic performance, which is automatically selected through the use of a rolling window approach. Four combination schemes are considered, two of which are simple averaging schemes in the form of the simple mean and trimmed mean, as literature shows that more complicated methods struggle to outperform these methods. In addition, a weighting scheme based on past performance introduced by (Stock & Watson, 2004) and a combination technique that uses regression based weights are assessed.

The results of the Monte Carlo simulations reveal that both bagging procedures can be utilised to substantially improve forecasting accuracy regardless of model specification. Similar observations are made for the application of these methods to monthly equity premium predictions. With the use of bagging procedures, the one-step-ahead monthly excess stock return forecasts based on macroeconomic variables outperform the historical average for short in-sample estimation periods. Although the application of forecast combination to bagged predictions is less effective than employing them to combine unbagged forecasts and combination does not yield improved forecasting accuracy compared to the ex-post best single model, the utilisation of the technique in tandem with bagging can still offer substantial benefits in practice. Specifically, the simultaneous use of both methods increases forecasting accuracy and reduces the accuracy risk that is related to model selection, as combining provides consistent performance over time compared to forecasts based on a single predictive variable. In contrast, the sequential implementation of bagging and subset combination using automatic selection, although less consistent than combining over the full set of predictors, provides more accurate predictions than the ex-post best single model in several cases.

The rest of this paper is organised into the following sections. Section 2 provides an overview of the relevant literature concerning equity premium prediction, bagging and forecast combination. Section 3 describes the models used to produce forecasts, the applied bagging procedures, forecast combination schemes and subset selection procedure. In addition, the performance measures used to evaluate forecasting accuracy are presented. This is followed by Section 4, which provides an overview of the DGPs employed in the Monte Carlo simulations followed by the results of these simulations. The practical application in the form of equity premium forecasting is discussed in Section 5. Finally, Section 6 summarises the main findings and supplies suggestions for further research.

2 Literature Review

Forecasting excess stock returns has been a central topic within the finance literature for decades. In the past few years, a considerable amount of papers concerning predictions based on financial and macroeconomic explanatory variables has accumulated. However, Welch and Goyal (2008) find that most variables examined in the literature fail to outperform the historical average out-of-sample. In contrast, Campbell and Thompson (2008) show that a number of predictive variables perform better than the historical average when some minor restrictions are imposed and suggest that even small gains in forecasting accuracy could be meaningful for investors in practice. The performance of numerous predictive variables and forecasting models has been investigated since, a number of papers consider bagging or forecast combination to improve prediction accuracy. Jin et al. (2014), Jordan, Vivian and Wohar (2017) and Hillebrand, Lukas and Wei (2021) make use of bagging in their equity premium forecasting procedure and find that some models achieve better results than the historical average when using this technique. Similarly, Rapach et al. (2010) and Tsiakas, Li and Zhang (2020) implement forecast combination and report significantly improved equity premium predictions.

Although bagging has recently seen considerable success in time series settings, it was first introduced by (Breiman, 1996) as a technique to reduce predictor variance for applications with IID data, which proved to be particularly effective for unstable predictors. Bühlmann and Yu (2002) demonstrate the effectiveness of bagging at decreasing variance and mean squared error for hard decision problems, which are inherently unstable. Even though bootstrap aggregating proved effective at increasing forecast accuracy in empirical applications, detailed understanding of the exact statistical workings of the technique was lacking. Friedman and Hall (2007) contribute to expanding this understanding by providing substantial statistical insight into the factors that affect the performance of bagging. In a similar vein, Stock and Watson (2012) show that bagging asymptotically is a shrinkage forecast. Lee and Yang (2006) extend the use of bagging to time series by employing it in binary and quantile prediction using asymmetric loss functions. Inoue and Kilian (2008) further explore this avenue through employing three variants of the technique to forecast the U.S. consumer price inflation. Jin et al. (2014) address the lack of statistical justification for the use of bagging in a time series context, by proving that their proposed bootstrap aggregating method, which is based on using historical forecasting information to improve predictions, reduces the MSPE. Petropoulos, Hyndman and Bergmeir (2018) further explore the causes behind the excellent forecasting performance that the technique provides when applied to time series and find that this accuracy mostly originates from reducing model uncertainty.

Unlike bagging, forecast combination has been used in time series applications since its introduction by (Bates & Granger, 1969). Combining forecasts is now common practice in the forecasting literature to improve the accuracy of time series predictions and it has become widely acknowledged that the technique brings substantial benefits (Wang, Hyndman, Li & Kang, 2023). Combining forecasts is often better than selecting a single best model, as the latter is unluckily to fully capture the complex characteristics that time series typically exhibit, such as time-varying trends, seasonality changes, and structural breaks (Clements & Hendry, 1998). In addition, Petropoulos et al. (2018) conclude that selecting a single model is subject to model, data and parameter uncertainty which can all be reduced through forecast combination. Although a vast number of refined combining methods have been proposed in the literature, Timmermann (2006) finds that it is challenging to beat simple, non-parametric, combination methods, partly because these do not suffer from estimation errors in determining the weights. Similarly, Makridakis, Spiliotis and Assimakopoulos (2020) report that simple combinations continue to perform relatively well compared to more complicated weighting schemes and machine learning algorithms. Other forecasting methods tend to be held back by the challenges of estimation. For example, the generalised version of the optimal weights proposed by (Bates & Granger, 1969) relies on accurately estimating the covariance matrix of the forecasts, which is notoriously difficult. Despite this, some more complicated methods have been implemented to realise forecasting performance gains, like the weights based on historic model performance proposed by (Stock & Watson, 2004).

Despite the fact that most combination approaches assessed in the literature use all available models in constructing combined predictions, the characteristics of the models used for combining have a major impact on forecasting performance, as including models with low accuracy will affect the final results negatively and a high degree of diversity among the included forecasts is instrumental to realising improved performance (Thomson, Pollock, Önkal & Gönül, 2019). The approach that logically follows from these findings is to include only a subset consisting of the most appropriate models in combined forecasts, which has displayed notable effectiveness in academic research. Zhou, Wu and Tang (2002) show that in a neural network application, selecting a subset of models before combining reduces variance as well as bias. Similarly, Hibon and Evgeniou (2005) and Lichtendahl and Winkler (2020) find that using only a subset of all available forecasts to create combination forecasts often yields favourable results as it may reduce accuracy risk. The application of subset combination might also produce positive outcomes for equity premium predictions, as Geweke and Amisano (2011) observe that combining many instead of all forecasts could lead to improved accuracy. Although subset combination has been shown to improve forecast accuracy, most of the considered methods are challenging to implement and suffer from limited portability to existing procedures, as they rely on ad-hoc modelling decisions. Therefore Kourentzes, Barrow and Petropoulos (2019) propose a heuristic to automatically identify a forecast subset that at least matches the performance of more involved methods In a similar vein, Diebold and Shin (2019) explore a direct subset combination procedure and find that it outperforms the simple averaging scheme and achieves comparable accuracy to the ex-post single best model.

The contribution of this paper to the literature is twofold. First, this study examines the potential greater forecasting accuracy offered by employing both bagging and forecast combination techniques in tandem to construct monthly on-step-ahead equity premium predictions. While previous studies have demonstrated the effectiveness of these methods individually, there is limited research covering a sequential approach using both techniques, especially concerning application to equity premium prediction. Second, a novel forecasting method that combines bagging and subset combination based on automatic rolling-window selection is presented, which is relatively simple to implement and outperforms the ex-post best model based on a single predictor in several instances.

3 Methodology

This section first describes the parametric and non-parametric forecasting models used in this paper. Following this, both bagging methods and the forecast combination schemes are presented. Finally, a description of the forecast evaluation metrics is given.

3.1 Forecasting Models

In this study, the forecasting performance of both parametric and non-parametric forecasting models is evaluated. This allows for the examination of the forecasting accuracy provided by bagging and forecast combinations for misspecified as well as correctly specified models. Specifically, the considered parametric model is misspecified for the Monte Carlo simulations and equity premium forecasting, in contrast, the non-parametric models theoretically converge to the true DGPs under certain conditions (Racine et al., 2008). For the sake of simplicity only linear models are taken into consideration.

For all models, to construct a prediction of the dependent variable y_{t+1} at time t, a training set

$$\mathcal{D}_t \equiv \{(y_i, x_{i-1})\}_{i=t-R+1}^{t} \quad \text{for } t = R, \ldots, T-1, \tag{1}$$

is used, consisting of R observations. Here x_t is a $q \times 1$ vector of explanatory variables. Using this training set $\mathcal{D}_t$ and the input vector x_t in one of the models discussed below results in the one-step-ahead forecast for y_{t+1}, which is denoted as $\phi(x_t, \mathcal{D}_t)$.

The first model to be examined is the simple linear model

$$y_{t+1} = \beta' x_t + \epsilon_{t+1}, \tag{2}$$

with β being the $q \times 1$ parameter vector. An estimate for β at time t can be obtained by applying ordinary least squares (OLS) using $\mathcal{D}_t$. This results in the estimate

$$\hat{\beta}_t \equiv \left(\sum_{i=t-R}^{t-1} x_i x_i' \right)^{-1} \left(\sum_{i=t-R}^{t-1} x_i y_i \right), \tag{3}$$

which is used at time t to construct the standard one-step-ahead forecast for y_{t+1} given by

$$\phi_1(x_t, \mathcal{D}_t) \equiv \hat{\beta}_t' x_t \quad \text{for } t = R, \ldots, T-1. \tag{4}$$

In addition to the parametric forecasting model, two non-parametric forecasting models are considered. These are of the form

$$y_{t+1} = m(x_t) + \epsilon_{t+1}, \tag{5}$$

where the functional form of the smoothing function $m(\cdot)$ is estimated from the data and x_t does not include a constant term.

The first non-parametric forecasting model is the one-step-ahead local constant forecast

$$\phi_2(x_t, \mathcal{D}_t) \equiv \frac{\sum_{i=t-R}^{t-1} y_{i+1} K_h(x_i - x_t)}{\sum_{i=t-R}^{t-1} K_h(x_i - x_t)} \quad \text{for } t = R, \ldots, T-1, \tag{6}$$

where $K_h(x_i - x_t) = \prod_{j=1}^{q} h_j^{-1} k\left(\frac{x_{ij}-x_{tj}}{h_j}\right)$ is the univariate kernel function, with x_{ij} being the $\mathrm{j_{th}}$ entry of x_i, and $h_1, \ldots, h_q$ the kernel smoothing parameter sequences. The latter converge to zero as $t \to \infty$, as described in (Ullah & Pagan, 1999). The standard normal kernel function, $k(x) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})$, is used for both the Monte Carlo simulations and empirical application, as in Jin et al. (2014). The value of the smoothing parameters will be determined according to the rule of thumb $h_j = c_0 \hat{\sigma}_j R^{-1/(4+q)}$, following Silverman (1986), where c_0 is set to 2[1] and $\hat{\sigma}_j$ is the sample standard deviation of x_{tj}. The same bandwidth is used for all bootstrap resamples. This rule of thumb is applied for the sake of simplicity, as (Jin et al., 2014) find that the results produced by the bagging methods are robust to the choice of smoothing parameter. It is however important to note that the bandwidth choice does affect the simple forecasts produced by both of the non-parametric models.

The second non-parametric forecast considered is the one-step-ahead local linear forecast

$$\phi_3(x_t, \mathcal{D}_t) \equiv e_1'(X_t K_t X_t')^{-1} X_t K_t Y_t \quad \text{for } t = R, \ldots, T-1 \tag{7}$$

with e_1 is a $(q+1) \times 1$ vector with the first entry equal to 1 and all the other entries equal to 0, $X_T = ((1, (x_{t-R} - x_t)')', \ldots, (1, (x_{t-1} - x_t)')')$, $K_t = \mathrm{diag}(K_h(x_{t-R} - x_t), \ldots, K_h(x_{t-1} - x_t))$ and $Y_t = (y_{t-R}, \ldots, y_{t-1})'$.

3.2 Bagging and Revised Bagging

The forecasts described in the previous section are all constructed using only the data in the observed training set $\mathcal{D}_t$ Assuming that the R observations in each $\mathcal{D}_t$ are drawn from a known distribution $\mathbf{P}$, an improved forecast can be constructed in the form of the ensemble aggregating predictor

$$\phi_A(x_t) \equiv \mathbb{E}_{\mathcal{D}_t}[\phi(x_t, \mathcal{D}_t)], \tag{8}$$

where $\mathbb{E}_{\mathcal{D}_t}[\cdot]$ is the expectation with respect to $\mathcal{D}_t$. Breiman (1996) proved that MSPE of $\phi_A(x_t)$ is smaller than or equal to that of $\phi(x_t, \mathcal{D}_t)$, under the assumption that $(y_{i+1}, x_i) \in \mathcal{D}_t$ for $i = t - R + 1, .., t$, are IID draws from $\mathbf{P}$. It goes without saying that this assumption does not hold for time series. To justify the use of bagging in a time series application, Jin et al. (2014) relaxed this IID condition and proved

$$\mathbb{E}[y_{t+1} - \phi\,(x_t, \mathcal{D}_t)]^2 \geq \mathbb{E}[y_{t+1} - \mathbb{E}[\phi(x_t, \mathcal{D}_t)|x_t, y_{t+1}]]^2. \tag{9}$$

The left-hand side of the equation is the MSPE of the original forecast and the right-hand side of the equation is formed by the MSPE of the new predictor $\mathbb{E}[\phi(x_t, \mathcal{D}_t)|x_t, y_{t+1}]$. It seems infeasible to construct a consistent estimator for the latter. Because at time t, y_{t+1} is unknown. However, Jin et al. (2014) show that this predictor can be approximated by $\mathbb{E}[\phi(x_t, \mathcal{D}_t)|x_t]$, for

[1] c_0 is set equal to 5 when producing forecasts for DGP 1, with $(\alpha, \beta) = (0.7, 0.2)$ in Equation 20, to avoid singularity problems in Equation 7, caused by sparse data.

which it is possible to construct a consistent estimator, under certain assumptions. Therefore, the original forecast $\phi(x_t, \mathcal{D}_t)$ can be improved upon by employing this new ensemble aggregating predictor. It is important to note that Jin et al. (2014) prove the forecasts produced by all models described in Section 3.1 to be suitable for this purpose under certain regularity conditions. This new predictor and the traditional forecast introduced by (Breiman, 1996) are particularly effective at increasing forecasting accuracy for unstable predictors, which is well suited to equity premium forecasting, as the relation between the excess stock returns and its predictive variables is highly unstable (Hillebrand et al., 2021).

To construct an ensemble aggregating predictor in practice, multiple training sets drawn from P are needed. Since in reality, the distribution P is unknown, these additional training sets are constructed using moving block bootstrap, which uses randomly drawn blocks of observations from $\mathcal{D}_t$ and combines these to form a new training set. To create one such additional training set $\mathcal{D}_t^\star$ the following procedure is used:

1. The optimal block length $\mathcal{B}$, as described by (Politis & White, 2004), is determined based on the correlation structure of the observations of $y_t \in \mathcal{D}_t$.
2. The number of blocks to be drawn, n, is set *s.t.* $(n-1) \times \mathcal{B} \leq R \wedge n \times \mathcal{B} \geq R$.
3. The length of all blocks is set to $\mathcal{B}$. The length of block n is changed to $\mathcal{B}^\star$, with $\mathcal{B}^\star$ *s.t.* $(n-1) \times \mathcal{B} + \mathcal{B}^\star = R$
4. The starting points of the blocks, expressed as indices of y_i, $i_1, \dots i_n$ are IID draws from the discrete uniform distribution on the domain $\{t-R+1, \dots, t-\mathcal{B}+1\}$.
5. The series that results from putting these blocks one after another forms $\mathcal{D}_t^\star$.

Repeating this process B times results in the series of training sets $\{\mathcal{D}_t^{\star(b)}\}_{b=1}^B$. Using the forecasts produced based on these gives the traditional bagging predictor

$$\phi_{it}^\star \equiv \phi_i^\star(x_t, \mathcal{D}_t) \equiv \sum_{b=1}^{B} w_{b,t}\phi_i(x_t, \mathcal{D}_t^{\star(b)}) \quad \text{for } i = 1, 2, 3, \tag{10}$$

where $w_{b,t}$ is the weight function with $\sum_{b=1}^B w_{b,t} = 1$.

Jin et al. (2014) propose an updated bagging predictor that utilises a 2-step procedure to produce improved forecasts by exploiting historical information. The first step incorporates the historical information through a non-parametric regression of $\{\phi_i(x_j, \mathcal{D}_j)\}_{j=t-\bar{R}+1}^t$ on x_t with $\bar{R}$ being the number of forecasts incorporated in the regression. This results in the predictor

$$\hat{\mathbb{E}}[\phi_i(x_t, \mathcal{D}_t)\,|x_t] \equiv \frac{\sum_{j=t-\bar{R}+1}^t \phi_{ij} K_h(x_j - x_t)}{\sum_{j=t-\bar{R}+1}^t K_h(x_j - x_t)} \quad \text{for } t = R+\bar{R}, \dots, T-1 \text{ and } i = 1, 2, 3, \tag{11}$$

which has been shown to be a consistent estimator for $\mathbb{E}[\phi_i(x_t, \mathcal{D}_t)|x_t]$ by (Jin et al., 2014), under the conditions that $R \to \infty$, $\bar{R} \to \infty$, $Rh_1 \cdot \ldots \cdot h_q \to \infty$, $\bar{R}h_1 \cdot \ldots \cdot h_q \to \infty$ and $\sum_{i=1}^q h_j^2 \to 0$ in addition to standard conditions on the DGP. In theory, this means that employing this predictor without bagging would produce a lower MSPE than the traditional forecast $\phi_i(x_t, \mathcal{D}_t)$.

However, in practice the dataset might only allow for moderately large values of R and $\bar{R}$. Furthermore, using only a single training sample causes a high degree of forecast uncertainty. Therefore (Jin et al., 2014) propose to apply bagging, which results in the predictor

$$\hat{\mathbb{E}}^{\star}\phi_{it} \equiv \sum_{b=1}^{B} w_{b,t}\hat{\mathbb{E}}[\phi_i(x_t, \mathcal{D}_t^{\star(b)}|x_t] \quad \text{for } t = R+\bar{R}, \dots, T-1 \text{ and } i = 1,2,3. \tag{12}$$

For this bagging predictor and the predictor in Equation 10 the weights are set $w_{b,t} = \frac{1}{B}$ for the sake of simplicity, which entails that every bootstrapped training set is treated as equally important regardless of in-sample performance. Jin et al. (2014) implement the same weighting scheme because other options, such as the Bayesian model averaging implemented by (Lee & Yang, 2006), produce comparable results.

3.3 Forecast Combination

Combining multiple forecasts constructed using different models often produces superior results compared to selecting a single best forecast, as a single specification is unlikely to fully capture the intricacies of an underlying unknown data generating process (Wang et al., 2023). Therefore this paper will investigate the performance of multiple linear combination forecasts in conjunction with bagging, these combination forecasts are of the form

$$\phi_{it,C} \equiv \sum_{j=1}^{N} w_{it,j}\phi_{it}(x_{t,j}, \mathcal{D}_{t,j}), \quad \text{for } i = 1,2,3, \tag{13}$$

where $x_{t,j}$ is the vector that contains the observations for the set of explanatory variables j at time t, $\mathcal{D}_{t,j} \equiv \{(y_i, x_{i-1,j})\}_{i=t-R+1}^{t}$, $w_{it,j}$ the weight corresponding to the forecast based on these and N the total number of explanatory variable sets.

The first two combining methods considered are the simple mean and trimmed mean. These have the advantage of being trivial to implement, do not suffer from estimation error and significantly reduce the bias and variance present in individual forecasts as described by (Palm & Zellner, 1992). In addition to this, the literature indicates that simple schemes often outperform more complicated methods (Timmermann, 2006). The simple mean is given by the weights $w_{it,j} = \frac{1}{N}$ and the trimmed mean can be constructed by setting the weight corresponding to the two most extreme forecasts to 0 and all other weights equal to $\frac{1}{N-2}$.

The third combining method, as outlined in (Stock & Watson, 2004), constructs combination forecasts using weights based on past model performance, which allows forecasters to take advantage of the varying predictive power of macroeconomic variables. Rapach et al. (2010) find that this technique significantly outperforms the historical average statistically and economically for a variety of out-of-sample periods. The weights are derived from historical model performance as follows,

$$w_{it,j} = \frac{m_{it,j}^{-1}}{\sum_{j=1}^{n} m_{it,j}^{-1}} \quad \text{for } i = 1,2,3, \tag{14}$$

with

$$m_{it,j} \equiv \sum_{s=t-\widetilde{R}}^{t-1} \theta^{t-s}(y_{s+1} - \phi_i(x_{s,j}, \mathcal{D}_{s,j}))^2 \quad \text{for } i = 1, 2, 3. \tag{15}$$

The historical performance of a model based on a specific set of explanatory variables is taken into account through the discounted MSPE, $m_{it,j}$. This value is determined using the $\widetilde{R}$ most recent observations and depends on the value of the discount factor θ. Following (Rapach et al., 2010), the values considered for θ are 1.0 and 0.9.

The fourth and final method utilises a nonnegativity-restricted least squares (NRLS) regression to determine the combination weights. Gunter (1992) finds that in an empirical setting, the combination forecasts produced using this method are at least as accurate and robust as those constructed using the simple mean. The weights for this method are determined by estimating them as the parameters in the regression

$$\begin{aligned} y_{s+1} = w_{si,1}\phi_i(x_{s,1}, \mathcal{D}_{s,1}) + \ldots + w_{si,n}\phi_i(x_{s,n}, \mathcal{D}_{s,n}) + \epsilon_{s+1}, \quad s.t.\ w_{si,1}, \ldots, w_{si,n} \geq 0 \\ \text{for } i = 1, 2, 3, \end{aligned} \tag{16}$$

with the observations and forecasts used for parameter estimation being those for $s = t - \widetilde{R}, \ldots, t-1$.

In addition to combined forecasts based on all N explanatory variable sets, combination predictions derived from a subset of the best performing predictive variables are also considered. Diebold and Shin (2019) show that such a subset combination procedure can outperform methods that make use of all available forecasts and match the performance of the ex-post most accurate model. To construct a subset of predictors for time $t = R + \bar{R} + \widetilde{R}, \ldots, T-1$ the explanatory variables are ranked based on their MSPE over the past $\widetilde{R}$ periods. The subset used for the combination forecasts consists of the n most accurate predictors with $3 \leq n \leq N$.

3.4 Forecast Evaluation

To assess the out-of-sample performance of all forecasting methods, two performance measures are employed. Firstly, the MSPE

$$\text{MSPE}(\phi_i) = \frac{1}{T - R - \bar{R}} \sum_{t=R+\bar{R}+\widetilde{R}}^{T-1} (y_{t+1} - \phi_{it})^2 \quad \text{for } i = 1, 2, 3, \tag{17}$$

is used to evaluate forecasting performance for the Monte Carlo simulations. To get the final result the MSPE is averaged over all replications. Secondly, the out-of-sample R-squared as suggested by (Campbell & Thompson, 2008)

$$R^2_{os} = 1 - \frac{\sum_{t=R+\bar{R}+\widetilde{R}}^{T-1}(y_{t+1} - \phi_{it})^2}{\sum_{t=R+\bar{R}+\widetilde{R}}^{T-1}(y_{t+1} - \bar{\phi})^2} \quad \text{for } i = 1, 2, 3, \tag{18}$$

is employed to quantify the forecasting performance in the empirical application of forecasting excess stock returns, with $\bar{\phi}$ being the average excess return over the past R periods. A positive R^2_{os} indicates that the equity premium forecasts produced by a model have a lower average

MSPE than the historical average and vice versa.

4 Monte Carlo Simulations

This section offers a description of the DGPs used for Monte Carlo simulations followed by the results of these simulations.

4.1 Data Generating Processes

The DGPs, which are based on the processes used by Jin et al. (2014),

$$\begin{aligned} &\text{DGP 1: } y_{t+1} = 0.95 y_t \exp(-y_t^2) + \varepsilon_{t+1} \\ &\text{DGP 2: } y_{t+1} = \frac{0.5}{1 + exp(-x_t)} + \varepsilon_{t+1}, \\ &\text{DGP 3: } y_{t+1} = 2\phi\,(x_t)\,x_t + \varepsilon_{t+1}, \end{aligned} \tag{19}$$

will be employed to construct the time series for the Monte Carlo simulations. In these DGPs $t = 0, .., T-1$, $\phi\,(\cdot)$ is the standard normal density function, x_t is constructed using the AR(1) process $x_t = \rho x_{t-1} + \epsilon_t$ with ρ being set to 0, 0.95 and ϵ_t is IID standard normally distributed. The error term ε_t follows the GARCH(1,1) process

$$\begin{aligned} \varepsilon_t &= v_t \eta_t \\ v_t^2 &= 1 + \alpha \varepsilon_{t-1}^2 + \beta v_{t-1}^2, \end{aligned} \tag{20}$$

where η_t is IID standard normally distributed and the values considered for (α, β) are (0,0), (0.3,0.4), (0.7,0.2).

The number of observations generated is such that out-of-sample period consists of 50 observations, as Jin et al. (2014) report that the chosen value does not have a considerable impact on the observed forecasting performance. The number of bootstrap resamples is set to $B = 100$ and 200 Monte Carlo replications are used.

4.2 Simulation Results

Tables 1, 2 and 3 contain the MSPE percentage gains with respect to the simple linear forecast in Equation 4. The forecasting methods are evaluated using four different choices of R, in $R = 20, 50, 100, 200$, and $\bar{R}$ is held constant at 20, as Jin et al. (2014) report that different values of $\bar{R}$ produce comparable outcomes. For the sake of brevity only the outcomes for $R = 20, 200$ are displayed in the previously mentioned tables, the figures for $R = 50, 100$ are available in Appendix A. The results were produced using parallel computing, specifically MATLAB's parallel computing toolbox[2], and all 12 cores of an i5-1240P CPU. The results can be summarised as follows.

First, all methods except for the local linear forecast without bagging outperform the simple linear benchmark in the vast majority of cases. The rationale behind this is twofold. The bandwidth resulting from the rule of thumb is sub-optimal for the local linear model which

[2]https://nl.mathworks.com/products/parallel-computing.html

Table 1: MSPE percentage gain with respect to the simple linear model: DGP 1

		(α, β)		
R	Forecast	(0,0)	(.3,.4)	(.7,.2)
20	$\phi_1^\star$	3.83	11.62	20.98
	$\hat{\mathbb{E}}^\star\phi_1$	4.02	11.72	21.37
	ϕ_2	3.41	3.58	16.41
	$\phi_2^\star$	3.72	11.73	21.39
	$\hat{\mathbb{E}}^\star\phi_2$	3.78	11.76	21.43
	ϕ_3	-8.77	-42.37	-14.94
	$\phi_3^\star$	4.28	11.15	20.74
	$\hat{\mathbb{E}}^\star\phi_3$	4.80	11.62	21.19
200	$\phi_1^\star$	-1.67	0.79	3.65
	$\hat{\mathbb{E}}^\star\phi_1$	-0.88	0.78	3.88
	ϕ_2	0.09	0.16	1.25
	$\phi_2^\star$	-1.62	0.79	3.90
	$\hat{\mathbb{E}}^\star\phi_2$	-0.95	0.80	4.02
	ϕ_3	-1.16	-11.66	-25.81
	$\phi_3^\star$	-1.78	0.64	3.05
	$\hat{\mathbb{E}}^\star\phi_3$	-0.88	0.76	3.82

This table displays the MSPE gains with respect to the simple linear model for DGP 1. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α, β) in Equation 20.

results in decreased forecasting accuracy. In addition, this model may yield very odd forecasts when there is an outlier at the forecast horizon which leads to a further increase in MSPE. The percentage gains for the methods that outperform the benchmark are generally larger when R is small (20, 50) and less sizeable for large R (100, 200). For example, all forecasting methods, except the local constant model without bagging, have a higher MSPE than the benchmark for DGP 1 with $(\alpha, \beta) =$ (0,0) when R is large. This is partly caused by the fact that the accuracy of the parameter estimates used in the benchmark model improves drastically as more observations become available for estimation.

Second, both the traditional bagging method and the revised version outperform the non-bagged simple linear, local constant and local linear models by a sizeable margin in most cases. For example, when $R = 20$ the MSPE reduction with respect to the simple linear forecast is higher than 6% with the exception of DGP 1 with $(\alpha, \beta) =$ (0,0). When $(\alpha, \beta) =$ (.7,.2), these accuracy gains increase to over 20% for DGP 1. For the same value of (α, β) and $\rho = 0.95$, DGP 2 and DGP 3 also show relatively high reductions in MSPE, with decreases of over 10% and 9% respectively. In addition, the bagged models provide similar increases in accuracy relative to the simple linear benchmark, even if the MSPEs of the non-bagged models differ drastically. Take as an example DGP 2 with $(\alpha, \beta) = (.7, .2)$, $\rho = 0.95$ and $R = 200$, the non-bagged MSPE percentage gains for the local constant and local linear model are 0.24% and -1.28%

Table 2: MSPE percentage gain with respect to the simple linear model: DGP 2

		(α,β)					
		(0,0)	(.3,.4)	(.7,.2)	(0,0)	(.3,.4)	(.7,.2)
R	Forecast		$\rho = 0$			$\rho = 0.95$	
20	$\phi_1^\star$	7.96	8.00	6.72	9.98	10.81	10.21
	$\hat{\mathbb{E}}^\star\phi_1$	8.08	8.12	6.84	10.21	10.91	10.35
	ϕ_2	3.36	3.67	2.62	6.66	6.26	6.51
	$\phi_2^\star$	7.98	7.94	6.69	10.02	10.85	10.23
	$\hat{\mathbb{E}}^\star\phi_2$	8.06	8.07	6.78	10.16	10.94	10.34
	ϕ_3	-8.88	-7.79	-6.79	-8.75	-9.43	-10.37
	$\phi_3^\star$	7.77	7.80	6.35	9.78	10.65	10.14
	$\hat{\mathbb{E}}^\star\phi_3$	7.96	8.05	6.68	10.09	10.74	10.35
200	$\phi_1^\star$	0.29	0.75	0.67	-2.11	-0.64	0.53
	$\hat{\mathbb{E}}^\star\phi_1$	0.31	0.75	0.68	-1.21	-0.32	0.54
	ϕ_2	-0.21	-0.01	0.19	-0.02	-0.09	0.24
	$\phi_2^\star$	0.29	0.75	0.66	-2.12	-0.63	0.48
	$\hat{\mathbb{E}}^\star\phi_2$	0.30	0.74	0.68	-1.40	-0.37	0.54
	ϕ_3	-1.40	-1.50	-0.91	-1.40	-1.75	-1.28
	$\phi_3^\star$	0.29	0.73	0.59	-2.10	-0.67	0.42
	$\hat{\mathbb{E}}^\star\phi_3$	0.32	0.73	0.65	-1.17	-0.33	0.50

This table displays the MSPE gains with respect to the simple linear model for DGP 2. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α,β) in Equation 20, followed by the values of ρ in the third row.

respectively, whereas all three models with revised bagging result in the relatively similar MSPE gains of 0.54%, 0.54% and 0.50%. This leads to the conclusion that the application of bagging to misspecified parametric models can be as effective as employing it to improve the accuracy of correctly specified non-parametric forecasting models.

Third, the revised bagging method outperforms the traditional bagging method for most parameter settings and in-sample estimation window lengths. For DGP 2 and 3 the performance of the revised method relative to the standard bagging procedure appears to depend on the value of ρ. Take as an example the MSPE gains of DGP 3 with $R = 200, \rho = 0$ and $(\alpha, \beta) = (.7, .2)$, which don't differ more than 0.01% between both methods compared to a difference that grows up to 9 times as large for $\rho = 0.95$. These characteristics of the revised bagging method are well suited to equity premium prediction as many of the predictive variables described in Section 5.1 are highly persistent.

5 Empirical Application: Equity Premium Prediction

This section details the application of bagging and forecast combination in tandem to monthly equity premium forecasting. First, an overview of the used data is given, followed by the outcomes of implementing bagging in isolation. Finally, the results for the utilisation of both techniques in sequence are presented.

5.1 Data

The empirical dataset containing the monthly equity premium values and macroeconomic explanatory variables runs from 1927 M1 to 2005 M12. The majority of data used to compute the equity premium and the predictive variables comes from the updated dataset of (Welch & Goyal, 2008) and can be retrieved from Prof. A Goyal's website [3]. The data for the real earnings and real prices were obtained from Prof. R. J. Shiller's website[4]. The equity premium is computed as the return on the S&P 500 index minus the risk-free rate which is derived from the Treasury-bill rate. It is important to note that the log of equity premium is used to construct the forecasts and the simple equity premium is employed to compute the R^2_{os}. The predictive variables used correspond to the ones employed by (Jin et al., 2014) and are Dividend Price Ratio (*d/p*), Earnings Price Ratio (*e/p*), Smoothed Earnings Price Ratio (*se/p*), Book-to-Market Ratio (*b/m*), Treasury Bill (*tbl*), Long Term Yield (*lty*), Term Spread (*ts*), Default Yield Spread (*ds*), Inflation (*inf*), Net Equity Expansion (*ntis*) and Lagged Equity Premium (*lagy*). A description of these macroeconomic variables can be found in Appendix B.

Table 3: MSPE percentage gain with respect to the simple linear model: DGP 3

		(α,β)					
		(0,0)	(.3,.4)	(.7,.2)	(0,0)	(.3,.4)	(.7,.2)
R	Forecast		$\rho = 0$			$\rho = 0.95$	
20	$\phi_1^\star$	7.57	7.60	7.83	10.26	10.64	9.02
	$\hat{\mathbb{E}}^\star\phi_1$	7.65	7.79	8.11	10.34	10.57	9.56
	ϕ_2	3.35	3.47	3.06	6.83	6.77	5.77
	$\phi_2^\star$	7.52	7.66	7.92	10.19	10.68	9.42
	$\hat{\mathbb{E}}^\star\phi_2$	7.54	7.72	8.09	10.26	10.62	9.56
	ϕ_3	-8.12	-7.49	-9.87	-8.79	-9.46	-5.30
	$\phi_3^\star$	7.54	7.48	7.58	10.21	10.55	8.90
	$\hat{\mathbb{E}}^\star\phi_3$	7.67	7.62	7.95	10.38	10.49	9.35
200	$\phi_1^\star$	0.51	0.59	0.41	-0.07	0.62	0.56
	$\hat{\mathbb{E}}^\star\phi_1$	0.55	0.60	0.42	0.09	0.64	0.63
	ϕ_2	-0.07	0.09	0.04	0.47	0.15	0.16
	$\phi_2^\star$	0.49	0.59	0.42	-0.14	0.57	0.54
	$\hat{\mathbb{E}}^\star\phi_2$	0.52	0.60	0.41	0.06	0.61	0.63
	ϕ_3	-1.26	-1.13	-1.33	-0.50	-1.55	-1.44
	$\phi_3^\star$	0.51	0.49	0.40	-0.10	0.61	0.53
	$\hat{\mathbb{E}}^\star\phi_3$	0.55	0.55	0.40	0.19	0.66	0.62

This table displays the MSPE gains with respect to the simple linear model for DGP 3. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α, β) in Equation 20, followed by the values of ρ in the third row.

[3] https://sites.google.com/view/agoyal145

[4] https://shillerdata.com/

5.2 Bagging Results

Table 4 contains the results of the application of bagging to equity premium forecasting. Following (Jin et al., 2014) the in-sample estimation window lengths considered are $R = 24, 60, 120$ and the number of observations used to estimate the revised bagging predictor is kept constant at $\bar{R} = 24$. The results can be summarised as follows.

First, none of the predictive variables is able to outperform the historical average forecast when used in the simple linear model for any of the tested values of R. From this, one might conclude that the macroeconomic variables do not have any predictive power for the excess returns on the S&P 500. However, the local linear model does outperform the historical average in some cases. For example, the Term Spread and Default Yield Spread provide a lower MSPE than the average excess return for all values of R. This leads to the conclusion that some of the macroeconomic variables provide a certain amount of information about the future equity premium. Even these small positive values of the R^2_{os}, like 0.98% for ds with $R = 120$, can be economically meaningful for investors (Campbell & Thompson, 2008).

Second, both bagging methods provide improved accuracy compared to their non-bagged counterparts for nearly every explanatory variable and value of R. These improvements relative to the non-bagged forecasts and historical average are especially sizeable for $R = 24$, here the R^2_{os} ranges from 2.75% to 4.10%. The gains over the historical average are, however, not limited to $R = 24$. For example, the local linear model, with $R = 120$, using the T-Bill rate gives a R^2_{os} of -5.71%. In contrast, applying the revised bagging method results in a value of 0.14%. A similar pattern emerges when comparing the accuracy of the revised bagging predictor to the traditional bagging method. The former outperforms the latter for the majority of explanatory variables and values of R. These improvements of the revised bagging method over the traditional are once again largest for $R = 24$. The values of the R^2_{os} for the revised range from 3.52% to 4.10% while those of traditional run from 2.75% to 3.71%.

5.3 Forecast Combination Results

The results of Equity Premium forecasting using forecast combination for the full set of variables as well as the subset combination are displayed in Table 5, 6 and 7. The settings considered for R and $\bar{R}$ are the same as described above. The values evaluated for the combination forecast hold-out period are $\widetilde{R} = 24, 60, 120$. In addition to the combination forecasts, the results of the models that only use a single predictor are also displayed for the same out-of-sample periods. For the subset combination methods an arbitrary and relatively small value of n was chosen, following (Diebold & Shin, 2019), in $n = 4$. For the sake of brevity the results for $n = 3, 5, 6$ are reported in Appendix C. The contents of these tables can be summarised as follows.

First, in general all combining methods, except the regression based weights, outperform a considerable number of the bagged models based on a single explanatory variable. However, the combined forecasts very rarely are more accurate than the best single bagged model. For example, in Table 7 for $\widetilde{R} = 120$, the 8 worst performing local linear forecasts with revised bagging give R^2_{os} values ranging from 0.01% to 0.29% and the simple mean produces a value of 0.31%, which is only bested by the 3 best performing single models. In contrast, the application of forecast combination to non-bagged models results in more accurate equity premium forecasts

Table 4: Out-of-Sample R squared (%): Monthly Equity Premium forecasts

		Variable										
R	Forecast	d/p	e/p	se/p	b/m	tbl	lty	ts	ds	inf	$ntis$	$lagy$
24	ϕ_1	-7.57	-9.13	-8.13	-11.05	-10.08	-8.21	-7.78	-6.23	-5.99	-5.61	-6.69
	$\phi_1^\star$	3.33	3.44	3.54	3.55	3.53	3.74	3.24	3.50	3.60	3.62	3.28
	$\hat{\mathbb{E}}^\star\phi_1$	3.95	3.67	3.84	3.78	3.81	4.01	3.71	3.89	3.67	4.10	3.73
	ϕ_2	-2.16	-0.77	-0.71	-3.51	0.47	0.28	0.28	1.41	-1.22	0.06	-2.76
	$\phi_2^\star$	3.52	3.42	3.53	3.50	3.63	3.61	3.56	3.43	3.40	3.48	3.53
	$\hat{\mathbb{E}}^\star\phi_2$	3.79	3.63	3.76	3.69	3.78	3.85	3.73	3.71	3.55	3.78	3.83
	ϕ_3	-13.88	-12.74	-12.51	-18.93	-15.77	-11.64	-11.89	-22.77	-8.69	-9.30	-13.16
	$\phi_3^\star$	3.38	4.03	3.67	3.43	3.59	3.70	3.24	3.54	3.71	3.60	2.75
	$\hat{\mathbb{E}}^\star\phi_3$	3.83	4.10	3.88	3.73	3.89	4.00	3.60	3.88	3.77	4.04	3.52
60	ϕ_1	-3.08	-3.73	-6.28	-6.85	-2.13	-5.68	-1.37	-2.11	-1.73	-1.64	-3.09
	$\phi_1^\star$	0.32	0.06	0.37	0.34	-0.11	0.11	-0.22	-0.11	-0.13	0.08	-0.50
	$\hat{\mathbb{E}}^\star\phi_1$	0.02	0.05	-0.21	-0.01	0.03	0.00	-0.01	-0.02	-0.17	-0.16	-0.08
	ϕ_2	-1.55	-2.17	-1.58	-4.07	-0.15	-0.77	0.73	0.14	-0.89	-0.40	-1.78
	$\phi_2^\star$	-0.05	-0.02	0.14	0.01	-0.01	0.15	-0.16	-0.13	-0.21	0.03	-0.23
	$\hat{\mathbb{E}}^\star\phi_2$	-0.17	-0.01	-0.10	-0.17	0.03	0.13	-0.12	-0.12	-0.24	-0.14	-0.13
	ϕ_3	-10.63	-8.59	-10.08	-13.01	-6.83	-8.32	-3.55	-5.07	-3.71	-5.69	-10.66
	$\phi_3^\star$	0.46	-0.01	0.24	0.43	-0.34	-0.10	-0.20	-0.24	-0.11	-0.04	-0.50
	$\hat{\mathbb{E}}^\star\phi_3$	0.24	0.06	-0.30	0.04	-0.06	-0.05	0.04	-0.24	-0.16	-0.30	-0.01
120	ϕ_1	-2.99	-2.21	-3.24	-3.62	-2.51	-2.84	-0.64	-0.69	-1.33	-1.96	-2.01
	$\phi_1^\star$	-0.19	-0.05	-0.26	-0.22	-0.13	-0.18	-0.20	-0.12	0.06	-0.11	-0.05
	$\hat{\mathbb{E}}^\star\phi_1$	-0.08	-0.13	-0.26	-0.18	0.25	0.13	0.25	-0.11	-0.06	-0.01	-0.05
	ϕ_2	-3.63	-3.03	-1.39	-4.10	-0.78	-0.79	0.28	0.98	-1.00	-0.70	-1.61
	$\phi_2^\star$	-0.14	-0.10	-0.12	-0.25	-0.05	0.05	-0.16	-0.16	0.03	-0.19	-0.18
	$\hat{\mathbb{E}}^\star\phi_2$	-0.08	-0.11	-0.07	-0.23	0.22	0.20	0.08	-0.10	-0.11	-0.11	-0.20
	ϕ_3	-11.95	-7.68	-6.58	-9.84	-5.71	-5.47	-1.66	-2.29	-3.43	-5.01	-6.94
	$\phi_3^\star$	-0.03	-0.06	-0.14	-0.08	-0.37	-0.44	-0.25	-0.22	0.18	-0.13	0.13
	$\hat{\mathbb{E}}^\star\phi_3$	0.07	-0.10	-0.19	-0.06	0.14	-0.13	0.27	-0.09	-0.12	0.00	0.01

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $R = 24, 60, 120$ start from 1954 M1, 1957 M1 and 1962 M1 respectively. The second row contains the employed predictors.

compared to the best single model in the vast majority of cases. Similarly, the average performance gains that combining provides relative to the corresponding single models are generally larger for the unbagged models, especially for the combined forecasts based on the whole set of predictors. For example, in Table 5, the average R^2os of the unbagged simple linear forecasts with $\widetilde{R} = 120$ is 8.47 percentage points lower than that of the simple mean, for the same model with revised bagging this difference is only 0.07. This overall lack in performance increase for the bagged forecasts can be explained by the fact that bagging and forecast combination improve forecasting accuracy in a somewhat similar manner, both methods reduce prediction variance by smoothing forecast instability, either through the inclusion of additional training samples or multiple predictive variables. These results however, do not necessarily lead to the conclusion that employing bagging and forecast combination in tandem can not offer an increase in forecasting performance in a practical application. Specifically, a considerable part of the performance of the best single model is based on ex-post knowledge. In other words, one would need to select the best single model before the out-of-sample period to realise the performance in practice. Selecting the best single model is done based on an in-sample fit measure or past forecasting performance, which does not guarantee that the picked model will produce the most accurate forecasts in the future relative to the other predictors. The figures in Appendix D, especially Figure 1, clearly show that the forecasting accuracy a model based on a single predictor provides compared to different models varies greatly over time. Although combining forecasts generally does not result in a MSPE reduction, Figure 2 shows that the forecasting performance of the simple mean is consistent over time relative to the models based on a single predictors, which can reduce accuracy risk in practical applications.

Second, the simple mean and trimmed mean combination forecast deliver the overall best performance and have comparable accuracy. In line with expectations, their performance is relatively consistent compared to the models that only use a single predictor. These combination methods are closely followed in accuracy by the discounted MSPE combination models, which generally produce comparable values of R^2_{os}. However, they are outperformed by the simple averaging schemes by a considerable margin in some cases. Take as an example the unbagged local linear forecasts for $\widetilde{R} = 120$ in Table 6, the trimmed mean gives a R^2_{os} of -1.96% and the discounted MSPE weights using $\theta = 0.9$ produce a value of -3.09%. The choice of θ appears to have a negligible impact on the accuracy of the combination method. The performance of the discounted MSPE models relative to the simple mean also seems to be independent of $\widetilde{R}$. For $\theta = 0.9$ this can be explained by the fact that earlier observations will have a negligible impact on the final weighting of the different single models. For $\theta = 1.0$ this pattern is more curious, it could be related to the choice of $\widetilde{R}$. A 2-year hold-out period might already be too long to fully take advantage of fluctuations in predictive power of the explanatory variables. In general, the combination forecasts based on regression weights perform poorly, especially for lower values of $\widetilde{R}$. For $\widetilde{R} = 24$ the NRLS weights produce values of R^2_{os} as low as -39.10%. This lack of accuracy is mostly caused by the estimation errors that come with determining the weights using a regression, estimation errors that the other methods do not suffer. Consistent with expectations, the weights can be estimated more accurately for higher values of $\widetilde{R}$. For example, in Table 5 for the local linear model with revised bagging and $\widetilde{R} = 120$, both regression based forecasts outperform all other combination methods with a $R^2_{os} = 3.95\%$

Third, the combination methods that use a subset of the available models to construct a forecast seem to provide very inconsistent performance relative to their full set counterparts, except for the regression based combinations. The NRLS combination method that does not use the full set of available forecasts consistently has higher accuracy than its full set counterpart, this is caused by the fact that most of the single variable forecasts are highly correlated. Therefore, including fewer models in the regression reduces the multicollinearity induced estimation errors, which in turn increases the forecasting performance. The results of subset selection, as stated before, are inconsistent compared to the non-subset combination methods for all other combination techniques. The accuracy is also highly dependent on the choice of n, which can be seen in Appendix C, and thus contains a considerable ex-post component. Nonetheless, subset selection can produce outstanding performance in some cases and not only outperform full set forecast combination but also the ex-post best single model. Take for example Table 5, for $\widetilde{R} = 24$ the subset trimmed mean based on the simple linear model with revised bagging gives an R^2os of 4.24% and the single models give values ranging from 3.60% to 4.11%.

Table 5: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 24$

		Variable											Combination					Subset Combination				
$\widetilde{R}$	Forecast	d/p	e/p	se/p	b/m	tbl	lty	tbs	ds	inf	$ntis$	$lagy$	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-7.08	-9.23	-7.74	-11.22	-10.39	-8.78	-7.95	-6.48	-5.87	-5.61	-6.84	0.66	0.67	0.43	0.19	-39.10	-3.69	-4.88	-4.00	-4.10	-22.42
	$\phi_1^\star$	3.29	3.41	3.52	3.46	3.35	3.61	3.11	3.43	3.47	3.61	3.22	3.48	3.47	3.48	3.48	-2.13	3.62	3.68	3.62	3.62	-2.09
	$\hat{\mathbb{E}}^\star\phi_1$	3.98	3.67	3.94	3.81	3.69	3.96	3.60	3.86	3.59	4.11	3.68	3.88	3.85	3.88	3.89	-3.23	4.23	4.24	4.23	4.23	-3.13
	ϕ_2	-2.19	-0.82	-0.53	-3.55	0.27	0.15	0.14	1.38	-1.27	0.07	-2.92	0.09	-0.06	0.08	0.04	-5.96	0.17	-0.59	0.14	0.09	-5.94
	$\phi_2^\star$	3.48	3.39	3.50	3.42	3.51	3.49	3.50	3.32	3.29	3.48	3.49	3.48	3.47	3.48	3.48	-2.05	3.48	3.50	3.48	3.48	-2.23
	$\hat{\mathbb{E}}^\star\phi_2$	3.79	3.63	3.75	3.66	3.69	3.77	3.69	3.63	3.45	3.80	3.80	3.73	3.72	3.73	3.73	-1.68	3.67	3.65	3.67	3.67	-1.28
	ϕ_3	-13.57	-13.00	-11.69	-18.88	-16.01	-12.36	-11.82	-23.53	-8.66	-9.38	-13.55	-1.28	-1.09	-1.65	-2.18	-37.63	-6.67	-7.49	-6.98	-7.32	-22.84
	$\phi_3^\star$	3.29	4.00	3.62	3.35	3.48	3.62	3.14	3.52	3.65	3.45	2.65	3.54	3.56	3.55	3.55	-1.62	3.67	3.72	3.67	3.68	-1.55
	$\hat{\mathbb{E}}^\star\phi_3$	3.82	4.10	3.92	3.77	3.83	3.99	3.51	3.89	3.73	3.95	3.44	3.90	3.88	3.90	3.91	-3.45	4.10	4.09	4.10	4.11	-3.17
60	ϕ_1	-7.23	-10.15	-7.85	-11.67	-11.53	-8.81	-8.61	-6.19	-6.09	-6.22	-6.69	-0.04	0.07	-0.14	-0.51	-5.75	-4.08	-4.31	-4.25	-4.65	-5.11
	$\phi_1^\star$	3.26	3.34	3.47	3.41	3.30	3.52	3.04	3.39	3.43	3.58	3.09	3.42	3.42	3.42	3.43	2.99	3.21	3.25	3.21	3.21	2.99
	$\hat{\mathbb{E}}^\star\phi_1$	3.97	3.74	3.88	3.74	3.44	3.69	3.38	3.84	3.52	3.93	3.57	3.77	3.75	3.77	3.77	1.64	3.89	3.94	3.89	3.89	1.60
	ϕ_2	-2.37	-1.29	-0.69	-3.82	-0.08	-0.32	-0.18	1.71	-1.28	-0.39	-3.01	-0.16	-0.29	-0.15	-0.22	0.95	-0.09	-0.36	-0.09	-0.23	1.09
	$\phi_2^\star$	3.47	3.39	3.47	3.38	3.46	3.47	3.44	3.30	3.24	3.39	3.39	3.43	3.43	3.43	3.43	3.24	3.29	3.30	3.29	3.29	3.26
	$\hat{\mathbb{E}}^\star\phi_2$	3.76	3.64	3.70	3.59	3.57	3.70	3.57	3.59	3.38	3.63	3.69	3.65	3.64	3.65	3.65	2.82	3.54	3.54	3.54	3.54	2.81
	ϕ_3	-14.20	-14.11	-12.08	-19.65	-17.39	-12.62	-12.50	-23.02	-8.25	-9.88	-13.00	-2.12	-1.83	-2.40	-3.09	-7.02	-8.08	-7.25	-8.17	-8.56	-6.84
	$\phi_3^\star$	3.20	3.73	3.54	3.27	3.45	3.54	3.10	3.37	3.65	3.44	2.55	3.46	3.48	3.46	3.46	2.78	3.47	3.43	3.47	3.47	2.78
	$\hat{\mathbb{E}}^\star\phi_3$	3.74	3.97	3.81	3.65	3.59	3.72	3.29	3.78	3.70	3.76	3.37	3.75	3.75	3.75	3.76	2.01	3.80	3.74	3.80	3.80	1.88
120	ϕ_1	-7.91	-10.64	-9.08	-11.57	-10.99	-9.86	-7.79	-6.26	-6.10	-6.38	-6.67	-0.01	0.08	-0.08	-0.52	-1.59	-4.09	-4.47	-4.17	-4.66	-1.24
	$\phi_1^\star$	3.25	3.31	3.44	3.47	3.34	3.42	3.06	3.41	3.40	3.63	3.09	3.43	3.42	3.43	3.43	3.90	3.19	3.23	3.19	3.19	3.90
	$\hat{\mathbb{E}}^\star\phi_1$	3.96	3.74	3.79	3.74	3.39	3.69	3.31	3.89	3.52	3.96	3.57	3.75	3.73	3.75	3.76	3.93	3.69	3.59	3.69	3.69	3.91
	ϕ_2	-2.59	-1.25	-1.09	-3.62	-0.25	-0.51	-0.23	1.92	-1.23	-0.51	-3.09	-0.21	-0.34	-0.19	-0.27	1.95	-0.17	-0.30	-0.16	-0.30	2.12
	$\phi_2^\star$	3.45	3.36	3.43	3.38	3.44	3.39	3.42	3.26	3.19	3.34	3.36	3.40	3.39	3.40	3.40	3.94	3.26	3.28	3.26	3.26	3.95
	$\hat{\mathbb{E}}^\star\phi_2$	3.76	3.63	3.66	3.58	3.55	3.67	3.55	3.57	3.38	3.59	3.67	3.63	3.63	3.63	3.63	3.97	3.52	3.52	3.52	3.52	3.96
	ϕ_3	-15.34	-14.99	-13.68	-18.13	-17.08	-14.01	-11.73	-24.51	-8.14	-9.80	-13.02	-2.08	-1.88	-2.28	-3.11	-2.07	-7.28	-5.43	-7.35	-7.72	-2.64
	$\phi_3^\star$	3.15	3.74	3.52	3.45	3.45	3.39	3.11	3.37	3.61	3.51	2.64	3.47	3.49	3.47	3.47	3.82	3.35	3.32	3.35	3.35	3.81
	$\hat{\mathbb{E}}^\star\phi_3$	3.70	4.00	3.72	3.70	3.50	3.70	3.20	3.83	3.70	3.81	3.40	3.75	3.74	3.75	3.75	3.95	3.68	3.59	3.68	3.68	3.95

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts with $R = 24$ and $n = 4$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1956 M1, 1959 M1 and 1964 M1 respectively. The first row shows what method was used to produce the forecasts: single variable, full set forecast combination or subset forecast combination. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DE and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

Table 6: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 60$

		Variable											Combination						Subset Combination					
$\widetilde{R}$	Forecast	*d/p*	*e/p*	*se/p*	*b/m*	*tbl*	*lty*	*tbs*	*ds*	*inf*	*ntis*	*lagy*	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-2.86	-3.13	-6.52	-7.02	-2.71	-5.13	-2.01	-1.55	-1.78	-2.31	-3.11	0.11	0.03	0.08	0.01	-33.75	-1.11	-1.14	-1.23	-1.36	-20.40
	$\phi_1^\star$	0.50	0.22	0.56	0.52	0.12	0.30	0.02	0.04	0.04	0.27	-0.37	0.23	0.24	0.23	0.23	-4.03	0.28	0.25	0.28	0.28	-4.02
	$\hat{\mathbb{E}}^\star\phi_1$	0.16	0.18	-0.08	0.12	0.12	0.11	0.13	0.17	-0.02	0.00	0.08	0.13	0.13	0.13	0.13	-6.07	-0.02	0.02	-0.02	-0.02	-6.09
	ϕ_2	-1.49	-2.07	-1.59	-4.15	-0.63	-0.83	0.23	0.80	-0.89	-0.85	-1.82	-0.50	-0.36	-0.49	-0.53	-15.49	-0.95	-0.77	-0.96	-1.03	-13.27
	$\phi_2^\star$	0.11	0.14	0.32	0.19	0.18	0.34	0.01	0.04	-0.05	0.16	-0.11	0.15	0.14	0.15	0.15	-3.35	0.13	0.12	0.13	0.13	-3.73
	$\hat{\mathbb{E}}^\star\phi_2$	-0.02	0.13	0.06	-0.02	0.16	0.28	0.02	0.06	-0.10	-0.03	0.00	0.07	0.08	0.07	0.07	-4.03	0.01	0.00	0.01	0.01	-3.81
	ϕ_3	-10.52	-8.32	-10.37	-13.33	-7.41	-7.86	-4.14	-2.61	-3.79	-5.95	-10.85	-2.03	-1.58	-2.24	-2.61	-37.50	-5.76	-4.72	-5.81	-6.06	-22.79
	$\phi_3^\star$	0.67	0.19	0.41	0.63	-0.11	0.09	0.03	-0.14	0.08	0.22	-0.35	0.22	0.26	0.22	0.22	-5.57	0.15	0.16	0.14	0.15	-5.70
	$\hat{\mathbb{E}}^\star\phi_3$	0.41	0.22	-0.17	0.19	0.04	0.02	0.19	-0.10	0.01	-0.10	0.18	0.15	0.15	0.15	0.15	-6.35	0.09	0.09	0.08	0.08	-5.88
60	ϕ_1	-2.94	-3.23	-6.66	-7.46	-3.01	-5.07	-2.27	-1.56	-1.84	-2.31	-3.19	-0.07	-0.15	-0.04	-0.13	-7.57	-1.42	-1.50	-1.45	-1.62	-7.03
	$\phi_1^\star$	0.40	0.11	0.48	0.43	0.00	0.17	-0.09	-0.09	-0.08	0.14	-0.47	0.12	0.13	0.12	0.12	-1.67	0.18	0.17	0.18	0.18	-1.67
	$\hat{\mathbb{E}}^\star\phi_1$	0.08	0.09	-0.16	0.02	-0.01	0.00	-0.01	0.09	-0.11	-0.08	-0.01	0.03	0.04	0.03	0.03	-2.43	0.04	0.09	0.04	0.04	-2.37
	ϕ_2	-1.58	-2.16	-1.62	-4.37	-0.85	-0.91	0.06	0.83	-0.91	-0.87	-1.86	-0.58	-0.43	-0.57	-0.62	-8.17	-0.64	-0.18	-0.65	-0.77	-7.71
	$\phi_2^\star$	0.02	0.03	0.25	0.10	0.07	0.23	-0.11	-0.07	-0.15	0.03	-0.21	0.04	0.04	0.04	0.04	-1.64	0.02	0.03	0.02	0.02	-1.53
	$\hat{\mathbb{E}}^\star\phi_2$	-0.10	0.03	-0.01	-0.10	0.06	0.18	-0.11	-0.03	-0.17	-0.13	-0.08	-0.02	-0.01	-0.02	-0.02	-2.11	0.02	0.04	0.02	0.02	-2.00
	ϕ_3	-10.85	-8.26	-10.52	-13.99	-7.64	-8.03	-4.42	-2.68	-3.84	-6.05	-11.19	-2.21	-1.77	-2.25	-2.77	-10.87	-3.47	-2.45	-3.61	-4.26	-9.82
	$\phi_3^\star$	0.54	0.08	0.29	0.52	-0.22	-0.01	-0.06	-0.28	-0.05	0.05	-0.50	0.10	0.13	0.10	0.10	-1.68	0.00	0.06	0.00	0.00	-1.68
	$\hat{\mathbb{E}}^\star\phi_3$	0.30	0.13	-0.28	0.08	-0.08	-0.06	0.05	-0.20	-0.09	-0.20	0.07	0.05	0.04	0.04	0.04	-2.11	-0.01	0.05	-0.02	-0.02	-2.24
120	ϕ_1	-3.12	-3.03	-7.19	-8.19	-2.91	-5.43	-2.00	-1.52	-1.96	-2.23	-3.43	-0.13	-0.19	-0.11	-0.19	-4.31	-1.34	-1.99	-1.35	-1.58	-4.01
	$\phi_1^\star$	0.51	0.19	0.53	0.57	0.09	0.27	0.02	-0.04	0.07	0.27	-0.40	0.22	0.23	0.22	0.22	-0.04	0.38	0.40	0.38	0.38	-0.04
	$\hat{\mathbb{E}}^\star\phi_1$	0.12	0.21	-0.21	0.05	0.07	0.03	0.10	0.22	-0.01	-0.02	0.09	0.10	0.11	0.10	0.10	-0.39	0.17	0.26	0.17	0.17	-0.35
	ϕ_2	-1.80	-2.13	-1.98	-4.78	-0.82	-1.21	0.29	0.91	-0.95	-0.94	-2.00	-0.67	-0.46	-0.66	-0.71	-5.68	-0.68	-0.18	-0.69	-0.81	-5.65
	$\phi_2^\star$	0.06	0.05	0.30	0.22	0.17	0.34	-0.02	-0.05	-0.03	0.12	-0.14	0.12	0.11	0.12	0.12	-0.08	0.06	0.07	0.06	0.06	-0.06
	$\hat{\mathbb{E}}^\star\phi_2$	-0.07	0.08	0.01	0.00	0.18	0.27	0.01	0.07	-0.08	-0.05	0.00	0.06	0.07	0.06	0.06	-0.06	0.04	0.07	0.04	0.04	-0.05
	ϕ_3	-11.82	-7.99	-11.43	-15.23	-7.76	-8.61	-4.03	-2.77	-4.16	-6.18	-11.79	-2.50	-1.96	-2.57	-3.09	-7.34	-4.91	-3.48	-4.96	-5.58	-7.69
	$\phi_3^\star$	0.68	0.24	0.34	0.59	-0.14	0.09	0.03	-0.21	0.10	0.16	-0.42	0.20	0.24	0.20	0.20	-0.03	0.20	0.28	0.20	0.20	-0.03
	$\hat{\mathbb{E}}^\star\phi_3$	0.39	0.34	-0.32	0.07	0.00	-0.05	0.16	-0.01	0.01	-0.14	0.20	0.13	0.13	0.13	0.13	-0.39	0.03	0.08	0.03	0.02	-0.35

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts with $R = 60$ and $n = 4$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1959 M1, 1962 M1 and 1967 M1 respectively. The first row shows what method was used to produce the forecasts: single variable, full set forecast combination or subset forecast combination. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DE and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

Table 7: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 120$

		Variable											Combination					Subset Combination				
$\widetilde{R}$	Forecast	*d/p*	*e/p*	*se/p*	*b/m*	*tbl*	*lty*	*tbs*	*ds*	*inf*	*ntis*	*lagy*	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-3.32	-2.48	-3.70	-4.05	-2.60	-3.04	-0.42	-0.72	-1.38	-1.61	-2.11	-0.27	-0.37	-0.30	-0.31	-25.48	-1.25	-0.77	-1.26	-1.30	-16.22
	$\phi_1^\star$	-0.15	0.01	-0.21	-0.19	-0.05	-0.12	-0.14	-0.09	0.11	-0.08	0.01	-0.07	-0.07	-0.07	-0.07	-5.04	-0.02	0.00	-0.02	-0.02	-4.66
	$\hat{\mathbb{E}}^\star\phi_1$	-0.06	-0.12	-0.28	-0.19	0.31	0.19	0.30	-0.09	-0.04	0.01	-0.02	0.04	0.04	0.04	0.04	-4.11	0.14	0.16	0.15	0.15	-4.10
	ϕ_2	-3.91	-3.25	-1.69	-4.41	-0.83	-0.88	0.45	1.03	-1.00	-0.75	-1.73	-0.80	-0.81	-0.87	-0.95	-12.03	-1.38	-1.76	-1.50	-1.63	-11.57
	$\phi_2^\star$	-0.11	-0.08	-0.08	-0.22	0.03	0.11	-0.10	-0.12	0.06	-0.19	-0.16	-0.06	-0.07	-0.06	-0.06	-3.55	-0.13	-0.14	-0.13	-0.13	-3.34
	$\hat{\mathbb{E}}^\star\phi_2$	-0.06	-0.11	-0.06	-0.22	0.28	0.26	0.13	-0.07	-0.09	-0.13	-0.20	0.00	0.00	0.00	0.00	-2.39	0.03	0.07	0.03	0.03	-3.37
	ϕ_3	-12.84	-8.18	-7.34	-10.59	-5.96	-5.83	-1.49	-2.41	-3.57	-4.82	-7.20	-2.68	-2.17	-3.07	-3.58	-34.19	-5.19	-4.59	-5.54	-6.15	-23.82
	$\phi_3^\star$	0.01	0.01	-0.10	-0.08	-0.28	-0.38	-0.19	-0.18	0.22	-0.08	0.22	-0.03	-0.02	-0.03	-0.03	-6.07	-0.14	-0.13	-0.14	-0.14	-5.65
	$\hat{\mathbb{E}}^\star\phi_3$	0.10	-0.06	-0.21	-0.08	0.20	-0.08	0.31	-0.07	-0.10	0.03	0.06	0.07	0.08	0.07	0.07	-4.45	0.11	0.15	0.11	0.11	-5.05
60	ϕ_1	-3.31	-2.51	-3.58	-4.15	-2.53	-3.16	-0.50	-0.85	-1.44	-1.26	-2.24	-0.38	-0.48	-0.38	-0.41	-12.96	-1.06	-0.66	-1.07	-1.15	-9.03
	$\phi_1^\star$	-0.03	0.13	-0.07	0.00	0.13	0.02	0.07	0.10	0.28	0.09	0.21	0.10	0.09	0.10	0.10	-1.55	0.09	0.08	0.09	0.09	-1.55
	$\hat{\mathbb{E}}^\star\phi_1$	0.04	0.02	-0.23	-0.10	0.34	0.25	0.35	0.06	0.13	0.24	0.16	0.15	0.15	0.15	0.15	-2.07	0.28	0.32	0.28	0.28	-2.07
	ϕ_2	-4.02	-3.29	-1.80	-4.59	-0.90	-0.94	0.36	1.01	-1.00	-0.69	-1.77	-0.87	-0.88	-0.88	-1.01	-10.90	-1.38	-1.78	-1.42	-1.64	-9.47
	$\phi_2^\star$	0.00	0.04	0.06	-0.08	0.16	0.24	0.08	0.05	0.23	-0.01	0.03	0.09	0.08	0.09	0.09	-1.62	0.03	0.06	0.03	0.03	-1.45
	$\hat{\mathbb{E}}^\star\phi_2$	0.03	0.02	0.03	-0.12	0.36	0.35	0.23	0.08	0.08	0.07	-0.01	0.12	0.12	0.12	0.12	-1.64	0.28	0.31	0.28	0.27	-1.56
	ϕ_3	-13.08	-8.33	-7.29	-10.86	-5.86	-6.13	-1.36	-2.44	-3.60	-4.63	-7.43	-2.84	-2.31	-2.91	-3.71	-15.11	-5.91	-5.37	-5.97	-6.82	-12.56
	$\phi_3^\star$	0.15	0.13	0.09	0.10	-0.07	-0.20	0.04	0.02	0.41	0.10	0.41	0.15	0.16	0.15	0.15	-2.56	0.15	0.16	0.15	0.15	-2.32
	$\hat{\mathbb{E}}^\star\phi_3$	0.21	0.08	-0.12	0.02	0.28	0.01	0.38	0.11	0.10	0.26	0.23	0.20	0.21	0.19	0.19	-2.53	0.22	0.26	0.21	0.22	-2.48
120	ϕ_1	-3.85	-2.81	-3.70	-4.04	-2.37	-2.39	-0.48	-0.85	-1.60	-1.24	-2.39	-0.47	-0.64	-0.47	-0.48	-9.45	-0.98	-1.28	-0.98	-0.98	-7.24
	$\phi_1^\star$	0.16	0.31	0.09	0.11	0.32	0.20	0.19	0.21	0.40	0.23	0.32	0.25	0.23	0.25	0.25	-0.99	0.15	0.13	0.15	0.15	-0.97
	$\hat{\mathbb{E}}^\star\phi_1$	0.09	0.13	-0.15	0.03	0.41	0.36	0.40	0.20	0.24	0.29	0.28	0.23	0.22	0.23	0.23	-1.00	0.38	0.34	0.38	0.38	-0.99
	ϕ_2	-4.44	-3.70	-1.94	-4.67	-0.90	-0.74	0.41	1.02	-1.18	-0.45	-1.94	-0.95	-0.98	-0.96	-1.10	-7.82	-1.82	-2.43	-1.84	-2.08	-7.74
	$\phi_2^\star$	0.09	0.15	0.15	-0.03	0.26	0.35	0.13	0.11	0.30	0.08	0.12	0.17	0.16	0.17	0.17	-1.07	0.05	0.08	0.05	0.05	-0.99
	$\hat{\mathbb{E}}^\star\phi_2$	0.07	0.11	0.08	-0.06	0.41	0.43	0.25	0.16	0.14	0.14	0.10	0.19	0.18	0.19	0.18	-0.99	0.26	0.30	0.26	0.26	-0.99
	ϕ_3	-12.95	-9.09	-7.26	-10.44	-5.56	-5.07	-1.15	-2.43	-3.86	-3.96	-8.17	-2.76	-2.26	-2.91	-3.72	-9.26	-6.17	-5.98	-6.28	-7.16	-9.30
	$\phi_3^\star$	0.35	0.30	0.29	0.25	0.17	-0.01	0.20	0.15	0.58	0.19	0.48	0.31	0.33	0.31	0.31	-1.60	0.28	0.28	0.28	0.28	-1.59
	$\hat{\mathbb{E}}^\star\phi_3$	0.27	0.22	0.01	0.16	0.42	0.18	0.48	0.29	0.23	0.29	0.32	0.31	0.31	0.31	0.31	-1.32	0.30	0.36	0.30	0.31	-1.22

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts with $R = 120$ and $n = 4$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1964 M1, 1967 M1 and 1972 M1 respectively. The first row shows what method was used to produce the forecasts: single variable, full set forecast combination or subset forecast combination. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DE and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

6 Conclusion

This paper introduces a forecasting method that sequentially applies bagging and forecast combination to improve the accuracy of time series predictions. The two bagging approaches considered are the traditional bagging method and the revised method proposed by (Jin et al., 2014). In addition, multiple combination weighting schemes are evaluated in the form the simple mean, trimmed mean, discounted MSPE based weights and regression based weights. This sequential technique is employed to produce equity premium forecasts based on macroeconomic variables to investigate whether applying bagging and forecast combination in tandem can improve the accuracy of monthly one-step-ahead equity premium predictions. Before applying both methods to excess return prediction, the forecasting performance of bagging is investigated through Monte Carlo simulations. Thereafter, both methods are employed in succession to construct equity premium forecasts, where combinations derived from the entire set of macroeconomic variables and an automatically selected subset of well performing predictors are considered.

The Monte Carlo simulations show that both bagging methods yield sizeable forecasting performance increases for all tested model specifications. The revised bagging method tends to outperform the traditional in the simulations. This higher accuracy is especially prolific for explanatory variables with high autocorrelation, which is useful for equity premium forecasting as many of the macroeconomic variables possess this property. When applied to monthly equity premium forecasting, although none of the macroeconomic variables consistently outperform the historical average when used in non bagged models, both of the bagging techniques provide higher accuracy than the historical average by a sizeable margin for short in-sample estimation periods. This shows that the equity premium is predictable using the information contained in the macroeconomic variables. The accuracy improvements of the equity premium predictions resulting from forecast combination relative to the models that only use a single predictor are comparably smaller for bagged forecasts. This however, does not mean that the application of bagging and forecast combination in tandem to monthly equity premium predictions does not hold any practical value. Using these techniques sequentially improves forecasting accuracy and reduces the accuracy risk related to forecast selection, as combining provides relatively consistent performance over time. In addition, although the combinations based on a subset of predictors are less consistent than their full set counterparts, subset combination in tandem with bagging has shown some potential, even besting the ex-post best single model in several cases.

Given the observed performance of the use of bagging and forecast combination in tandem for equity premium forecasting, it might be interesting to investigate their sequential use for other time series. Specifically, considering this method for forecasting macroeconomic variables could be valuable, as they possess many of the same properties as the excess returns in the form of structural instability and numerous possible predictors. It may also be worthwhile to add to the existing literature, by evaluating the performance provided through the sequential application of both methods using a purely statistical approach. In addition, considering the sizeable reduction in MSPE that subset combination provides in certain instances, it may be beneficial to further refine the subset selection procedure. The approach implemented in this paper is relatively crude and leaves room for improvement, such as automatic selection of the number of forecasts to include in the combination or more appropriate selection criteria.

References

Bates, J. M. & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140.

Bühlmann, P. & Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, *30*(4), 927–961.

Campbell, J. Y. & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Clements, M. & Hendry, D. F. (1998). *Forecasting economic time series.* Cambridge University Press.

Diebold, F. X. & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, *35*(4), 1679–1691.

Friedman, J. H. & Hall, P. (2007). On bagging and nonlinear estimation. *Journal of Statistical Planning and Inference*, *137*(3), 669–683.

Geweke, J. & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.

Gunter, S. I. (1992). Nonnegativity restricted least squares combinations. *International Journal of Forecasting*, *8*(1), 45–59.

Hibon, M. & Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of forecasting*, *21*(1), 15–24.

Hillebrand, E., Lukas, M. & Wei, W. (2021). Bagging weak predictors. *International Journal of Forecasting*, *37*(1), 237–254.

Inoue, A. & Kilian, L. (2008). How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, *103*(482), 511–522.

Jin, S., Su, L. & Ullah, A. (2014). Robustify financial time series forecasting with bagging. *Econometric Reviews*, *33*(5-6), 575–605.

Jordan, S. J., Vivian, A. & Wohar, M. E. (2017). Forecasting market returns: bagging or combining? *International Journal of Forecasting*, *33*(1), 102–120.

Kourentzes, N., Barrow, D. & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, *209*, 226–235.

Lee, T.-H. & Yang, Y. (2006). Bagging binary and quantile predictors for time series. *Journal of Econometrics*, *135*(1-2), 465–497.

Lichtendahl, K. C. & Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, *36*(1), 142-149. (M4 Competition)

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.

Palm, F. C. & Zellner, A. (1992). To combine or not to combine? issues of combining forecasts. *Journal of Forecasting*, *11*(8), 687–701.

Petropoulos, F., Hyndman, R. J. & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational*

Research, *268*(2), 545–554.

Politis, D. N. & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70.

Racine, J. S. et al. (2008). Nonparametric econometrics: A primer. *Foundations and Trends® in Econometrics*, *3*(1), 1–88.

Rapach, D. E. & Strauss, J. K. (2010). Bagging or combining (or both)? an analysis based on forecasting us employment growth. *Econometric Reviews*, *29*(5-6), 511–533.

Rapach, D. E., Strauss, J. K. & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821–862.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

Stock, J. H. & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting*, *23*(6), 405–430.

Stock, J. H. & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, *30*(4), 481–493.

Thomson, M. E., Pollock, A. C., Önkal, D. & Gönül, M. S. (2019). Combining forecasts: Performance and coherence. *International Journal of Forecasting*, *35*(2), 474–484.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, *1*, 135–196.

Tsiakas, I., Li, J. & Zhang, H. (2020). Equity premium prediction and the state of the economy. *Journal of Empirical Finance*, *58*, 75–95.

Ullah, A. & Pagan, A. (1999). *Nonparametric econometrics.* Cambridge university press Cambridge.

Wang, X., Hyndman, R. J., Li, F. & Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, *39*(4), 1518–1547.

Welch, I. & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Zhou, Z.-H., Wu, J. & Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial Intelligence*, *137*(1-2), 239–263.

A Monte Carlo Simulations Results

Tables 8, 9 and 10 contain the MSPE percentage gains with respect to the simple linear forecast in Equation 4 for the Monte Carlo simulations with $R = 50, 100$ and $\bar{R} = 20$.

Table 8: MSPE percentage gain with respect to the simple linear model: DGP 1

		(α, β)		
R	Forecast	(0,0)	(.3,.4)	(.7,.2)
50	$\phi_1^\star$	1.17	2.89	6.72
	$\hat{\mathbb{E}}^\star\phi_1$	1.83	2.93	6.94
	ϕ_2	1.09	0.80	6.63
	$\phi_2^\star$	1.17	2.85	7.05
	$\hat{\mathbb{E}}^\star\phi_2$	1.58	2.93	7.06
	ϕ_3	-5.08	-35.22	-13.92
	$\phi_3^\star$	1.18	2.55	6.41
	$\hat{\mathbb{E}}^\star\phi_3$	1.82	2.90	6.74
100	$\phi_1^\star$	-0.88	0.86	5.21
	$\hat{\mathbb{E}}^\star\phi_1$	-0.17	0.88	5.26
	ϕ_2	0.42	-0.34	2.71
	$\phi_2^\star$	-0.97	0.86	5.41
	$\hat{\mathbb{E}}^\star\phi_2$	-0.42	0.93	5.44
	ϕ_3	-2.02	-17.93	-22.31
	$\phi_3^\star$	-0.87	0.64	4.72
	$\hat{\mathbb{E}}^\star\phi_3$	-0.08	0.83	5.01

This table displays the MSPE gains with respect to the simple linear model for DGP 1. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α, β) in Equation 20.

Table 9: MSPE percentage gain with respect to the simple linear model: DGP 2

		(α,β)					
		(0,0)	(.3,.4)	(.7,.2)	(0,0)	(.3,.4)	(.7,.2)
R	Forecast		$\rho = 0$			$\rho = 0.95$	
50	$\phi_1^\star$	2.59	3.01	3.23	1.71	3.17	2.77
	$\hat{\mathbb{E}}^\star\phi_1$	2.58	3.03	3.27	2.27	3.28	2.97
	ϕ_2	0.79	1.00	1.21	1.54	1.24	1.18
	$\phi_2^\star$	2.51	2.95	3.32	1.67	3.18	2.79
	$\hat{\mathbb{E}}^\star\phi_2$	2.49	2.99	3.35	2.02	3.28	2.93
	ϕ_3	-3.71	-4.83	-2.98	-4.68	-5.57	-3.10
	$\phi_3^\star$	2.64	2.84	2.93	1.79	3.08	2.61
	$\hat{\mathbb{E}}^\star\phi_3$	2.62	2.96	3.03	2.36	3.22	2.89
100	$\phi_1^\star$	1.19	0.99	1.58	-0.99	0.92	1.09
	$\hat{\mathbb{E}}^\star\phi_1$	1.18	1.03	1.60	-0.31	1.09	1.19
	ϕ_2	-0.11	0.19	0.39	0.16	0.50	0.53
	$\phi_2^\star$	1.17	1.00	1.62	-1.05	0.92	1.10
	$\hat{\mathbb{E}}^\star\phi_2$	1.16	1.02	1.66	-0.54	1.04	1.19
	ϕ_3	-2.53	-2.90	-2.75	-3.20	-2.56	-2.92
	$\phi_3^\star$	1.17	0.96	1.39	-0.98	0.86	1.06
	$\hat{\mathbb{E}}^\star\phi_3$	1.17	1.00	1.46	-0.29	1.06	1.18

This table displays the MSPE gains with respect to the simple linear model for DGP 2. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α, β) in Equation 20, followed by the values of ρ in the third row.

Table 10: MSPE percentage gain with respect to the simple linear model: DGP 3

		(α,β)					
		(0,0)	(.3,.4)	(.7,.2)	(0,0)	(.3,.4)	(.7,.2)
R	Forecast		$\rho=0$			$\rho=0.95$	
50	$\phi_1^\star$	2.66	2.56	2.73	3.12	2.77	3.80
	$\hat{\mathbb{E}}^\star\phi_1$	2.65	2.60	2.76	3.19	2.75	3.75
	ϕ_2	0.87	0.72	0.84	2.04	1.53	1.02
	$\phi_2^\star$	2.61	2.58	2.75	3.07	2.74	3.81
	$\hat{\mathbb{E}}^\star\phi_2$	2.62	2.59	2.81	3.15	2.75	3.79
	ϕ_3	-3.05	-3.98	-3.27	-4.05	-4.28	-6.75
	$\phi_3^\star$	2.51	2.42	2.47	3.05	2.67	3.69
	$\hat{\mathbb{E}}^\star\phi_3$	2.54	2.50	2.65	3.23	2.71	3.70
100	$\phi_1^\star$	1.22	1.02	0.74	1.20	1.80	1.08
	$\hat{\mathbb{E}}^\star\phi_1$	1.24	1.01	0.81	1.31	1.78	1.24
	ϕ_2	0.40	-0.09	0.10	1.12	0.74	0.88
	$\phi_2^\star$	1.23	0.92	0.82	1.10	1.79	1.05
	$\hat{\mathbb{E}}^\star\phi_2$	1.24	0.91	0.85	1.21	1.78	1.34
	ϕ_3	-2.23	-2.97	-2.63	-1.77	-2.09	-1.62
	$\phi_3^\star$	1.14	1.09	0.63	1.30	1.81	1.43
	$\hat{\mathbb{E}}^\star\phi_3$	1.17	1.07	0.69	1.48	1.78	1.53

This table displays the MSPE gains with respect to the simple linear model for DGP 3. The first column shows the values of R that were tested, followed by the forecasting model used in the second column. The second row contains the values of (α,β) in Equation 20, followed by the values of ρ in the third row.

B Macroeconomic Variables

A brief description of the macroeconomic variables is given below.

- Dividend Price Ratio (d/p): This ratio compares dividends paid on the S&P 500 index to the price level. It is calculated as the difference between the log of dividends and the log of prices. The value of dividends is computed as a 12-month moving sum.

- Earnings Price Ratio (e/p): This ratio compares earnings on the S&P 500 index to the price level. It is calculated as the difference between the log of earnings and the log of prices. The value of earnings is computed as a 12-month moving sum.

- Smoothed Earnings Price Ratio (se/p): This ratio uses a 10-year moving average of real earnings divided by current real prices.

- Book-to-Market Ratio (b/m): This ratio is constructed by dividing the book value by the market value for the Dow Jones Industrial Average.

- Treasury Bill (tbl): The interest rate on a three month Treasury Bill.

- Long Term Yield (lty): The yield on long-term government bonds.

- Term Spread (ts): The term spread is the difference between long-term government bond yields and Treasury bill rates.

- Default Yield Spread (ds): This spread is computed as the difference between BAA and AAA-rated corporate bond yields.

- Inflation (inf): The inflation is Consumer Price Index for all urban consumers. This variable is lagged to avoid using future information, as the values become available with a delay.

- Net Equity Expansion ($ntis$): This ratio is computed by dividing the net issues by S&P listed stocks, for which a 12-month moving sum is used, by the total end-of-year market capitalisation of S&P stocks.

- Lagged Equity Premium ($lagy$): The value of excess stock returns lagged one-period.

C Subset Combination

Tables 11, 12 and 13 contain the results for the application of subset forecast combination to equity premium forecasting with $n = 3, 5, 6$.

Table 11: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 24$

		$n=3$					$n=5$					$n=6$				
$\widetilde{R}$	Forecast	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-5.74	-5.98	-6.06	-6.17	-18.59	-1.88	-2.46	-2.21	-2.41	-29.83	-1.88	-2.46	-2.21	-2.41	-29.83
	$\phi_1^\star$	3.63	3.76	3.63	3.63	-2.10	3.58	3.62	3.58	3.58	-2.14	3.58	3.62	3.58	3.58	-2.14
	$\hat{\mathbb{E}}^\star\phi_1$	4.38	4.33	4.38	4.38	-3.19	4.14	4.10	4.15	4.15	-3.12	4.14	4.10	4.15	4.15	-3.12
	ϕ_2	-0.14	-1.31	-0.17	-0.22	-5.93	0.03	-0.37	0.02	-0.02	-6.03	0.03	-0.37	0.02	-0.02	-6.03
	$\phi_2^\star$	3.44	3.50	3.44	3.44	-2.02	3.47	3.48	3.47	3.47	-2.06	3.47	3.48	3.47	3.47	-2.06
	$\hat{\mathbb{E}}^\star\phi_2$	3.75	3.69	3.75	3.75	-1.66	3.72	3.71	3.72	3.72	-1.68	3.72	3.71	3.72	3.72	-1.68
	ϕ_3	-8.13	-7.60	-8.56	-9.04	-22.65	-4.30	-4.96	-4.65	-5.02	-32.10	-4.30	-4.96	-4.65	-5.02	-32.10
	$\phi_3^\star$	3.80	3.94	3.80	3.81	-1.51	3.68	3.69	3.68	3.68	-1.62	3.68	3.69	3.68	3.68	-1.62
	$\hat{\mathbb{E}}^\star\phi_3$	4.09	4.05	4.09	4.09	-3.18	4.05	4.05	4.05	4.05	-3.24	4.05	4.05	4.05	4.05	-3.24
60	ϕ_1	-7.94	-7.23	-8.04	-8.26	-5.13	-3.93	-4.28	-4.06	-4.39	-5.11	-3.93	-4.28	-4.06	-4.39	-5.11
	$\phi_1^\star$	3.24	3.24	3.24	3.24	2.99	3.21	3.23	3.21	3.21	2.99	3.21	3.23	3.21	3.21	2.99
	$\hat{\mathbb{E}}^\star\phi_1$	3.88	3.94	3.88	3.88	1.68	3.87	3.87	3.87	3.88	1.56	3.87	3.87	3.87	3.88	1.56
	ϕ_2	-0.37	-0.33	-0.36	-0.53	1.13	-0.07	-0.24	-0.06	-0.18	1.02	-0.07	-0.24	-0.06	-0.18	1.02
	$\phi_2^\star$	3.23	3.26	3.22	3.23	3.22	3.32	3.32	3.32	3.32	3.25	3.32	3.32	3.32	3.32	3.25
	$\hat{\mathbb{E}}^\star\phi_2$	3.51	3.52	3.51	3.51	2.82	3.60	3.62	3.59	3.60	2.81	3.60	3.62	3.59	3.60	2.81
	ϕ_3	-11.55	-8.95	-11.67	-11.93	-7.71	-6.25	-5.91	-6.35	-6.81	-6.97	-6.25	-5.91	-6.35	-6.81	-6.97
	$\phi_3^\star$	3.46	3.53	3.46	3.46	2.80	3.43	3.45	3.43	3.43	2.78	3.43	3.45	3.43	3.43	2.78
	$\hat{\mathbb{E}}^\star\phi_3$	4.00	3.98	4.00	4.00	1.97	3.86	3.83	3.86	3.87	2.01	3.86	3.83	3.86	3.87	2.01
120	ϕ_1	-5.69	-5.17	-5.76	-6.34	-1.26	-2.83	-2.57	-2.91	-3.38	-1.14	-2.83	-2.57	-2.91	-3.38	-1.14
	$\phi_1^\star$	3.14	3.17	3.14	3.14	3.89	3.24	3.29	3.24	3.24	3.90	3.24	3.29	3.24	3.24	3.90
	$\hat{\mathbb{E}}^\star\phi_1$	3.78	3.72	3.78	3.78	3.92	3.69	3.61	3.70	3.70	3.92	3.69	3.61	3.70	3.70	3.92
	ϕ_2	-0.26	0.01	-0.25	-0.45	2.13	-0.11	-0.33	-0.10	-0.21	2.05	-0.11	-0.33	-0.10	-0.21	2.05
	$\phi_2^\star$	3.27	3.22	3.27	3.27	3.94	3.30	3.32	3.30	3.30	3.95	3.30	3.32	3.30	3.30	3.95
	$\hat{\mathbb{E}}^\star\phi_2$	3.50	3.51	3.50	3.50	3.96	3.55	3.54	3.55	3.55	3.97	3.55	3.54	3.55	3.55	3.97
	ϕ_3	-7.72	-5.81	-7.82	-8.45	-2.10	-4.95	-3.65	-5.05	-5.43	-2.84	-4.95	-3.65	-5.05	-5.43	-2.84
	$\phi_3^\star$	3.28	3.24	3.28	3.28	3.81	3.37	3.34	3.37	3.37	3.81	3.37	3.34	3.37	3.37	3.81
	$\hat{\mathbb{E}}^\star\phi_3$	3.62	3.50	3.62	3.63	3.93	3.67	3.64	3.67	3.68	3.95	3.67	3.64	3.67	3.68	3.95

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts using subset combination with $R = 24$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1956 M1, 1959 M1 and 1964 M1 respectively. The first row shows the values of n evaluated. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DE and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

Table 12: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 60$

		$n=3$					$n=5$					$n=6$				
$\widetilde{R}$	Forecast	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-2.40	-3.84	-2.51	-2.61	-17.46	-1.09	-1.05	-1.15	-1.21	-23.16	-1.09	-1.05	-1.15	-1.21	-23.16
	$\phi_1^\star$	0.31	0.20	0.31	0.31	-4.03	0.24	0.22	0.24	0.24	-4.02	0.24	0.22	0.24	0.24	-4.02
	$\hat{\mathbb{E}}^\star\phi_1$	0.03	0.10	0.03	0.02	-6.08	-0.04	0.02	-0.04	-0.04	-6.17	-0.04	0.02	-0.04	-0.04	-6.17
	ϕ_2	0.01	0.05	-0.06	-0.13	-10.73	-0.61	-0.44	-0.63	-0.70	-14.66	-0.61	-0.44	-0.63	-0.70	-14.66
	$\phi_2^\star$	0.12	0.15	0.12	0.12	-3.69	0.13	0.13	0.13	0.13	-3.85	0.13	0.13	0.13	0.13	-3.85
	$\hat{\mathbb{E}}^\star\phi_2$	-0.01	-0.09	-0.01	-0.01	-3.67	0.00	0.00	0.00	0.00	-3.70	0.00	0.00	0.00	0.00	-3.70
	ϕ_3	-6.90	-7.43	-6.87	-6.98	-21.95	-4.37	-3.85	-4.59	-4.99	-23.64	-4.37	-3.85	-4.59	-4.99	-23.64
	$\phi_3^\star$	0.09	0.03	0.09	0.10	-5.30	0.17	0.20	0.17	0.17	-5.68	0.17	0.20	0.17	0.17	-5.68
	$\hat{\mathbb{E}}^\star\phi_3$	0.00	0.03	0.00	-0.01	-5.85	0.06	0.11	0.06	0.06	-5.91	0.06	0.11	0.06	0.06	-5.91
60	ϕ_1	-1.82	-2.86	-1.87	-2.03	-6.66	-0.51	-0.49	-0.55	-0.68	-6.65	-0.51	-0.49	-0.55	-0.68	-6.65
	$\phi_1^\star$	0.22	0.15	0.22	0.22	-1.67	0.19	0.17	0.19	0.19	-1.67	0.19	0.17	0.19	0.19	-1.67
	$\hat{\mathbb{E}}^\star\phi_1$	-0.08	-0.05	-0.08	-0.08	-2.35	0.00	0.03	0.00	0.00	-2.37	0.00	0.03	0.00	0.00	-2.37
	ϕ_2	-0.62	-0.11	-0.64	-0.74	-8.44	-0.53	-0.17	-0.55	-0.65	-7.97	-0.53	-0.17	-0.55	-0.65	-7.97
	$\phi_2^\star$	0.10	0.10	0.10	0.10	-1.52	0.06	0.03	0.06	0.06	-1.55	0.06	0.03	0.06	0.06	-1.55
	$\hat{\mathbb{E}}^\star\phi_2$	0.02	-0.06	0.02	0.02	-1.99	0.00	0.02	0.00	0.00	-1.97	0.00	0.02	0.00	0.00	-1.97
	ϕ_3	-6.19	-5.64	-6.29	-6.67	-10.54	-2.51	-1.36	-2.65	-3.18	-9.98	-2.51	-1.36	-2.65	-3.18	-9.98
	$\phi_3^\star$	0.02	0.10	0.02	0.02	-1.68	0.08	0.10	0.08	0.08	-1.68	0.08	0.10	0.08	0.08	-1.68
	$\hat{\mathbb{E}}^\star\phi_3$	-0.09	0.13	-0.09	-0.09	-2.39	0.01	0.05	0.01	0.01	-2.11	0.01	0.05	0.01	0.01	-2.11
120	ϕ_1	-2.20	-3.77	-2.21	-2.53	-4.02	-1.05	-1.69	-1.06	-1.20	-3.79	-1.05	-1.69	-1.06	-1.20	-3.79
	$\phi_1^\star$	0.44	0.45	0.44	0.44	-0.04	0.37	0.37	0.37	0.37	-0.04	0.37	0.37	0.37	0.37	-0.04
	$\hat{\mathbb{E}}^\star\phi_1$	0.08	0.18	0.08	0.07	-0.34	0.11	0.17	0.11	0.11	-0.44	0.11	0.17	0.11	0.11	-0.44
	ϕ_2	-1.26	-0.33	-1.27	-1.40	-6.07	-0.57	0.06	-0.58	-0.68	-5.67	-0.57	0.06	-0.58	-0.68	-5.67
	$\phi_2^\star$	0.05	-0.01	0.05	0.05	-0.06	0.08	0.06	0.08	0.08	-0.06	0.08	0.06	0.08	0.08	-0.06
	$\hat{\mathbb{E}}^\star\phi_2$	0.05	-0.03	0.05	0.05	-0.05	0.02	0.04	0.02	0.02	-0.05	0.02	0.04	0.02	0.02	-0.05
	ϕ_3	-7.68	-5.80	-7.72	-8.26	-8.36	-4.61	-3.24	-4.64	-5.12	-7.43	-4.61	-3.24	-4.64	-5.12	-7.43
	$\phi_3^\star$	0.11	0.15	0.11	0.11	-0.03	0.19	0.21	0.19	0.19	-0.03	0.19	0.21	0.19	0.19	-0.03
	$\hat{\mathbb{E}}^\star\phi_3$	-0.07	0.05	-0.07	-0.07	-0.39	0.06	0.10	0.06	0.06	-0.38	0.06	0.10	0.06	0.06	-0.38

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts using subset combination with $R = 60$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1959 M1, 1962 M1 and 1967 M1 respectively. The first row shows the values of n evaluated. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DE and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

Table 13: Out-of-Sample R squared: Monthly Equity Premium single and combined forecasts $R = 120$

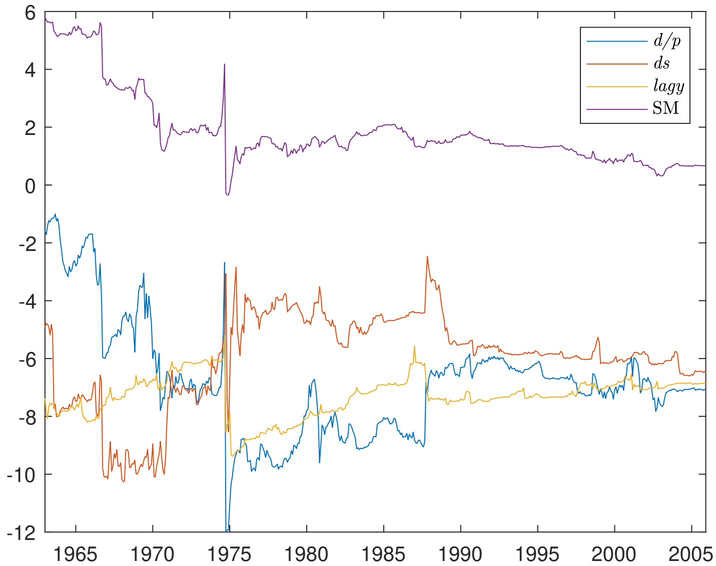
		$n=3$					$n=5$					$n=6$				
$\widetilde{R}$	Forecast	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS	SM	TM	DE(1.0)	DE(0.9)	NRLS
24	ϕ_1	-1.80	-0.87	-1.79	-1.83	-15.12	-1.12	-1.25	-1.13	-1.15	-16.28	-1.12	-1.25	-1.13	-1.15	-16.28
	$\phi_1^\star$	-0.08	-0.06	-0.08	-0.08	-4.65	-0.03	-0.02	-0.03	-0.03	-4.69	-0.03	-0.02	-0.03	-0.03	-4.69
	$\hat{\mathbb{E}}^\star\phi_1$	0.16	0.16	0.16	0.16	-4.32	0.14	0.13	0.14	0.14	-4.48	0.14	0.13	0.14	0.14	-4.48
	ϕ_2	-1.89	-3.55	-1.99	-2.09	-11.90	-1.05	-1.07	-1.18	-1.32	-11.63	-1.05	-1.07	-1.18	-1.32	-11.63
	$\phi_2^\star$	-0.16	-0.16	-0.16	-0.16	-3.49	-0.11	-0.13	-0.11	-0.11	-3.37	-0.11	-0.13	-0.11	-0.11	-3.37
	$\hat{\mathbb{E}}^\star\phi_2$	0.06	0.15	0.06	0.06	-3.12	0.03	0.05	0.03	0.03	-3.01	0.03	0.05	0.03	0.03	-3.01
	ϕ_3	-6.99	-8.18	-7.10	-7.49	-23.46	-4.33	-4.03	-4.65	-5.20	-23.93	-4.33	-4.03	-4.65	-5.20	-23.93
	$\phi_3^\star$	-0.15	-0.13	-0.15	-0.15	-5.40	-0.11	-0.13	-0.11	-0.11	-5.87	-0.11	-0.13	-0.11	-0.11	-5.87
	$\hat{\mathbb{E}}^\star\phi_3$	0.02	0.09	0.02	0.02	-5.01	0.07	0.08	0.07	0.08	-5.07	0.07	0.08	0.07	0.08	-5.07
60	ϕ_1	-1.13	-0.84	-1.14	-1.22	-9.05	-0.99	-0.66	-1.00	-1.06	-8.99	-0.99	-0.66	-1.00	-1.06	-8.99
	$\phi_1^\star$	0.02	0.04	0.02	0.02	-1.56	0.11	0.08	0.11	0.11	-1.55	0.11	0.08	0.11	0.11	-1.55
	$\hat{\mathbb{E}}^\star\phi_1$	0.35	0.30	0.35	0.35	-2.08	0.25	0.26	0.25	0.25	-2.07	0.25	0.26	0.25	0.25	-2.07
	ϕ_2	-2.38	-4.24	-2.40	-2.60	-9.64	-1.15	-1.23	-1.19	-1.40	-9.55	-1.15	-1.23	-1.19	-1.40	-9.55
	$\phi_2^\star$	-0.06	-0.01	-0.06	-0.06	-1.49	0.04	0.07	0.04	0.04	-1.55	0.04	0.07	0.04	0.04	-1.55
	$\hat{\mathbb{E}}^\star\phi_2$	0.29	0.33	0.29	0.29	-1.47	0.23	0.28	0.23	0.23	-1.59	0.23	0.28	0.23	0.23	-1.59
	ϕ_3	-5.56	-3.47	-5.71	-6.66	-12.55	-4.83	-4.03	-4.93	-5.87	-14.77	-4.83	-4.03	-4.93	-5.87	-14.77
	$\phi_3^\star$	0.11	0.25	0.11	0.10	-1.88	0.16	0.16	0.16	0.16	-2.54	0.16	0.16	0.16	0.16	-2.54
	$\hat{\mathbb{E}}^\star\phi_3$	0.23	0.18	0.23	0.23	-2.47	0.17	0.23	0.17	0.18	-2.48	0.17	0.23	0.17	0.18	-2.48
120	ϕ_1	-0.48	-0.76	-0.49	-0.51	-6.58	-0.44	-0.78	-0.45	-0.47	-6.94	-0.44	-0.78	-0.45	-0.47	-6.94
	$\phi_1^\star$	0.16	0.11	0.16	0.16	-0.97	0.16	0.13	0.16	0.16	-0.99	0.16	0.13	0.16	0.16	-0.99
	$\hat{\mathbb{E}}^\star\phi_1$	0.48	0.46	0.48	0.48	-0.98	0.34	0.30	0.34	0.34	-0.99	0.34	0.30	0.34	0.34	-0.99
	ϕ_2	-2.36	-4.50	-2.37	-2.56	-7.88	-1.35	-1.60	-1.37	-1.62	-7.71	-1.35	-1.60	-1.37	-1.62	-7.71
	$\phi_2^\star$	0.04	0.10	0.04	0.04	-0.98	0.12	0.13	0.12	0.12	-1.05	0.12	0.13	0.12	0.12	-1.05
	$\hat{\mathbb{E}}^\star\phi_2$	0.37	0.46	0.37	0.37	-0.89	0.24	0.24	0.24	0.24	-0.99	0.24	0.24	0.24	0.24	-0.99
	ϕ_3	-8.24	-11.06	-8.30	-8.75	-9.01	-6.02	-5.74	-6.09	-6.78	-9.10	-6.02	-5.74	-6.09	-6.78	-9.10
	$\phi_3^\star$	0.24	0.25	0.24	0.24	-1.51	0.25	0.28	0.25	0.25	-1.59	0.25	0.28	0.25	0.25	-1.59
	$\hat{\mathbb{E}}^\star\phi_3$	0.40	0.43	0.40	0.40	-1.23	0.23	0.24	0.23	0.23	-1.30	0.23	0.24	0.23	0.23	-1.30

This table displays the R^2_{os} (%) for the Monthly Equity Premium forecasts using subset combination with $R = 120$. The first column shows the values of $\widetilde{R}$ that were tested, followed by the forecasting model used in the second column. The out-of-sample periods for $\widetilde{R} = 24, 60, 120$ start from 1964 M1, 1967 M1 and 1972 M1 respectively. The first row shows the values of n evaluated. The second row displays the employed predictors or forecast combination weighting scheme. For the latter SM, TM, DD and NRLS denote simple mean, trimmed mean, discounted MSPE weights followed by the value of θ and regression based weights respectively.

D Cumulative Out-Of-Sample R-Squared

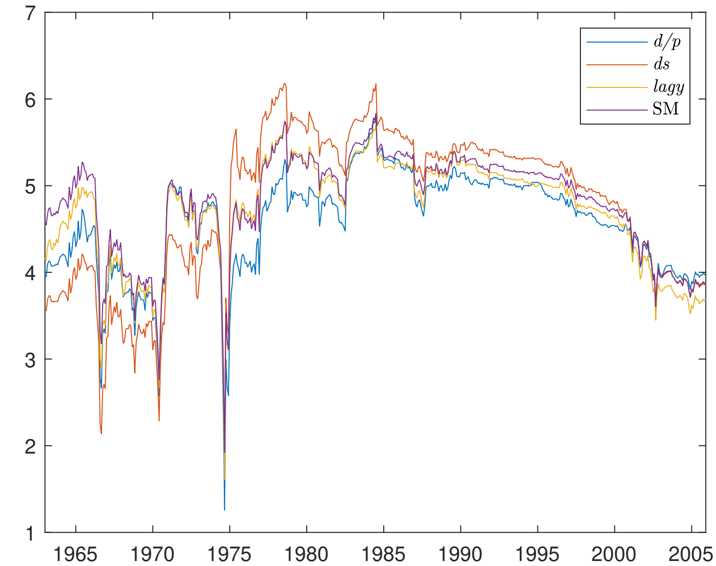
Figure 1 and 2 display the R^2_{os} at each point in time (cumulative R^2_{os}) for forecasts produced using the simple linear model and the simple mean combined forecasts, with $R = \widetilde{R} = 24$.

Figure 1: Cumulative R^2_{os} non-bagged



This figure displays the cumulative R^2_{os} for the non-bagged forecasts produced using the simple linear model based on the Dividend Price Ratio, Default Yield Spread and Lagged Equity Premium as well as the simple mean combined forecasts for the full set of variables. The value employed for R and $\widetilde{R}$ is 24 and the displayed period runs from 1963 M1 to 2005 M12.

Figure 2: Cumulative R^2_{os} bagged



This figure displays the cumulative R^2_{os} for the bagged forecasts produced using the simple linear model based on the Dividend Price Ratio, Default Yield Spread and Lagged Equity Premium as well as the simple mean combined forecasts for the full set of variables. The value employed for R and $\widetilde{R}$ is 24 and the displayed period runs from 1963 M1 to 2005 M12.

E Code

This appendix lists all MATLAB code and data files used to compute the results reported in this paper. A concise description of each file is provided.

- The excel file **Raw Data** contains the original data.
- All computed macroeconomic variables and final results are stored in the excel file **Variables and Results**.
- All data generated by the DGPs described in 4.1 and forecasts computed for the Monte Carlo Simulations as well as equity premium predictions are stored in the following MATLAB data files.
  - **Results_MC.mat** contains the data and forecasts for the Monte Carlo simulations.
  - **Results_EP.mat** contains the non-bagged and bagged equity premium forecasts.
  - **Results_EP_Combined.mat** contains the combined premium forecasts.
- To replicate the results presented in this paper, the following MATLAB scripts should be run in order.
  - **Monte_Carlo_main.m** runs the Monte Carlo simulations to produce the results reported in Table 1, 2, 3, 8, 9 and 10.
  - **Generate_Variables.m** computes the equity premium and agronomic variables according to the description provided in Section 5.1 and Appendix B.
  - **EP_Forecast_main.m** constructs the non-bagged and bagged equity premium predictions to produce the results reported in Table 4.
  - **EP_Forecast_Combined_main.m** constructs combined equity premium predictions to produce the results reported in Table 5, 6 and 7. The figures in Table 11, 12 and 13 can be replicated by changing the value of n.
  - **Generate_Figs.m** produces Figure 1 and 2.
- The following helper functions are called in some of the previous scripts.
  - **MC.m** runs a Monte Carlo simulation for the chosen DGP and parameter settings and handles parallelisation.
  - **MonteCarloIter.m** runs one replication of a Monte Carlo simulation.
  - **DGP1.m**, **DGP2.m** and **DGP3.m** generate a time series according to DGP 1, DGP 2 and DGP 3 in Equation 19 respectively.
  - **EPForecast.m** produces non-bagged and bagged equity premium predictions for specific values of R and $\bar{R}$.
  - **OLS.m**, **LocalConstant.m** and **LocalLinear.m** produce one-step-ahead predictions according to Equation 4, 6 and 7 respectively.

- **KernelBandwidth.m** computes the appropriate value for the smoothing parameter according to the rule-of-thumb described in Section 3.1.
- **TrimmedMean.m**, **MSPEWeighted.m** and **NRLS.m** construct combined forecasts according to their respective procedures as described in 3.3.
- **BlockBootstrap.m** produces B bootstrap resamples according to the procedure described in Section 3.2.
- **opt_block_length_REV_dec07.m** computes the optimal bootstrap block length as described by (Politis & White, 2004), the code can be retrieved from Prof. A. Patton's website[5].
- **mlag.m** generates a matrix of n lags from a matrix containing a set of vectors, the code can be retrieved from Prof. A. Patton's website.
- **R2OOSLog.m** calculates the R^2_{os} according to Equation 18.
- **Rank.m** returns a matrix containing the forecasts of the n best performing models based on a single predictor according to the MSPE as described in Section 3.3.

[5]https://public.econ.duke.edu/ ap172/code.html