Predicting the Equity Premium with Combination Forecasts: A Reappraisal*

Sebastian Denk and Gunter Löffler†

January 15, 2022

The present paper reappraises the usefulness of forecast combination for predicting the US equity premium. For comparison, we also include penalized regression and dimension reduction approaches. We fail to find evidence of predictive ability in recent decades, regardless of the forecasting method used. The decline in return predictability is accompanied by a decline in the ability to predict the real economy.

JEL classification: C58, G12, G17

Keywords: Equity premium, return predictability, forecast combination

*We thank Carol Alexander for helpful comments, and Hai Lin for providing us with the code for the iterated combination forecasts.

†Ulm University. Contact: gunter.loeffler@uni-ulm.de

1 Introduction

Forecasting stock returns is a challenging task. Welch and Goyal (2008) showed that a large number of variables suggested in the literature fail to predict the equity premium in an out-of-sample setting. Using the same set of variables, however, Rapach, Strauss, and Zhou (2010) arrived at a positive assessment of predictability: Combining individual forecasts from predictors studied by Welch and Goyal (2008) generates out-of-sample forecasts that are more accurate than the historical average benchmark forecast.

In the present paper, we evaluate whether combining forecasts still helps to predict the equity premium. Such a review is timely. As shown by Goyal, Welch, and Zafirov (2021), who do not examine combination forecasts, a large number of predictive models that produced superior out-of-sample forecasts in the original papers did not deliver in recent years. Apparently, the standard out-of-sample methodology provides little assurance that predictability will persist. One reason for the instability could be data snooping. According to Rytchkov and Zhong (2020) and Dichtl et al. (2021), it could also be relevant for the interpretation of the results in Rapach, Strauss, and Zhou (2010).

In our study, we do not only update combination forecasts in the way they were implemented by Rapach, Strauss, and Zhou (2010). Motivated by recent contributions to the predictability literature, we also consider variations of the combination forecast method as well as alternatives for dealing with a large number of predictors.

As variations, we consider using weighted least squares instead of ordinary least squares for the individual forecast regressions (Johnson (2018)); iterated combination forecasts (Lin, Wu, and Zhou (2018)); and the combination of forecasts obtained after applying an elastic net (Rapach and Zhou (2020)). As alternative statistical procedures, we consider ridge, LASSO and elastic net regressions (Li and Tsiakas (2017)), principal components (Rapach and Zhou (2013)), and three-pass regression filters (Kelly and Pruitt (2015) and Rytchkov and Zhong (2020)).

To alleviate possible concerns that conclusions are based on an update period that is relatively short compared to the original sample period, we proceed as follows: As in Rapach, Strauss, and Zhou (2010), we make the first out-of-sample forecast for 1965. The last forecast is made for the end of 2020. We then split this period into two equal-sized sub-periods, one from 1965 to 1992 and another more recent one, from 1993 to 2020.

Our results can be summarized as follows: Until the early 1990s, most of the forecast methods we examine show a forecast quality that is significantly higher than the one of the historical mean benchmark. Since then, however, none of the methods shows significant out-of-sample forecasting gains relative to the historical mean. Results do not depend on whether we judge forecast gains based on relative predictive accuracy, or based on utility gains of a mean-variance investor who uses the forecast rather than the historical mean to determine the optimal equity allocation.

Rapach, Strauss, and Zhou (2010) argue that the ability of combining methods to predict returns can be rationalized by the observation that combination forecasts are linked to the real economy. We study whether the strength of these links has changed and find a decline in the ability of combination forecasts to predict industrial production growth and real GDP growth. One possible explanation for the decline in the quality of the equity premium forecasts is therefore that the links to the real economy have faded, making it more difficult to track fluctuations in the risk premium associated with the business cycle. Another explanation could be that the forecast gains reported in Rapach, Strauss, and Zhou (2010) are spurious, consistent with the conclusions of Rytchkov and Zhong (2020) and Dichtl et al. (2021). Note that a spuriousness of published results does not require any data snooping activity from the side of Rapach, Strauss, and Zhou (2010). As shown by Rytchkov and Zhong (2020), using predictors that were suggested in prior literature leads to data-snooping biases in forecast methods that aggregate the information contained in these predictors. This does not conclude the list of possible explanations for the results we find. The predictability in the early years could have been based on market inefficiencies that later disappeared. Finally, the decline in predictability that we observe could be transitory.

Judging which explanation contributes the most is very difficult but it is also not relevant for the main conclusion of our analysis. The key result of Rapach, Strauss, and Zhou (2010) is that "economic variables collectively are valuable and consistently outperform the historical average forecast of the equity premium" (p. 823). While this characterization is correct for the time period from the 1960s to the early 1990s, we document that it no longer applies to a more recent period of equal length.

Several closely related papers on the robustness of equity premium predictability were already mentioned. Farmer, Schmidt, and Timmermann (2019) and Harvey et al. (2021) identify pockets of predictability associated with some of the variables also used in the combination forecasts. Since they document such pockets also for recent time periods, their work does not provide direct support to the possible hypothesis that the decline in the usefulness of combination is transitory. Pastor and Stambaugh (2001) and Lettau and Van Nieuwerburgh (2008) identify shifts in the equity premium and financial ratios, respectively. Such shifts can make it harder to predict the equity premium, but it is not obvious that they undermine the usefulness of combination forecasts because combining can be seen as a way of dealing with structural instability (cf. Hendry and Clements (2004) and Rapach, Strauss, and Zhou (2010)).

We proceed as follows. In Section 2, we set the stage by describing the different forecasting strategies. The methodology for evaluating forecasts is sketched in Section 3. Empirical results are presented in Section 4. Section 5 concludes.

2 Forecasting strategies

2.1 The standard predictive model

The workhorse of the literature on equity premium predictability is a linear regression model:

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

where r_{t+1} denotes the equity premium, i.e., the return on a stock market index at time $t+1$ minus the return on a risk-free asset. Typical candidates for $x_{i,t}$, a predictor variable observable at time t, include financial ratios, interest rates, and macroeconomic variables. Most studies use ordinary least squares (OLS) to estimate the parameters α and β. In our analysis, we also consider weighted least squares (WLS). For the WLS regressions, we follow Johnson (2018) by setting weights equal to the inverse of the conditional variance of equity premia.

This paper focuses on out-of-sample tests of return predictability. In such tests, only information that would have been available at time t is used to estimate the parameters of equation (1). An out-of-sample forecast $\hat{r}_{i,t+1}$ that is made for time $t+1$ using predictor i has the following form:

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

where $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$ are the coefficient estimates and their time index indicates that these parameters are estimated recursively, using only information that was available at time t.

Extending equation (1) to a multivariate setting is straightforward. A model in which all predictor candidates are incorporated without applying a model selection process is usually called a kitchen sink model.

2.2 Forecast combination

Combining methods from Rapach, Strauss, and Zhou (2010)

A combination forecast of the next-period equity premium made at time t is a weighted average of N individual forecasts:

$$\hat{r}_{t+1}^{C} = \sum_{i=1}^{N} \omega_{i,t} \hat{r}_{i,t+1}, \tag{3}$$

where $\omega_{i,t}$ is the weight on the forecast generated with variable i at time t, with $\sum_{i=1}^{N} \omega_{i,t} = 1$. The time subscript of $\omega_{i,t}$ indicates that the weights can only be chosen based on information available at time t. This does not preclude cases in which the weights are constant.

In our analysis, we implement all averaging rules used in Rapach, Strauss, and Zhou (2010). The mean combination assigns each individual forecast an equal weight $\omega_{i,t} = 1/N$ in equation (3). The median combination corresponds to the median of the individual return forecasts. For the trimmed mean combination forecast, we discard the individual forecasts with the largest and smallest return prediction and set $w_{i,t} = 1/(N-2)$ for the forecasts that we keep for combining.

These simple averaging rules do not consider the track record of individual forecasts. A combining method that puts greater weight on better-performing individual forecasts is the discounted mean squared prediction error (DMSPE) forecast, where the weights of a given individual forecast depend inversely on its mean squared prediction error. The weights of the DMSPE forecast are given as[1]

$$\omega_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{i=1}^{N} \phi_{i,t}^{-1}}, \tag{4}$$

where

[1] Note that the DMSPE forecast requires a hold-out period before the out-of-sample evaluation period to estimate of the combining weights. In the empirical application, we follow Rapach, Strauss, and Zhou (2010) in setting the length of the initial hold-out period to ten years.

$$\phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s}(r_{s+1} - \hat{r}_{i,s+1})^2. \tag{5}$$

Here, m denotes the last observation before the hold-out out-of-sample period that is used to calibrate the weights and θ is a discount factor. When $\theta < 1$, forecast accuracy in the more distant past is discounted relative to recent performance, with smaller values of θ indicating greater discounting. In addition to the values of 1.0 and 0.9 that are considered for θ in Rapach, Strauss, and Zhou (2010), we compute the DMSPE forecast for $\theta = 0.5$ to investigate whether greater discounting can improve predictive accuracy.

The iterated combination method

Lin, Wu, and Zhou (2018) suggest an iterated combination approach in which a simple combination forecast is additionally combined with the historical average of the equity premium. The simple combination can for example be based on the mean or median of the individual forecasts as described in the previous section. The epithet *iterated* is a reference to this additional step of applying a further combination to a combination forecast.

As laid out in Lin, Wu, and Zhou (2018), the theoretically optimal iterated forecast that minimizes the expected mean squared error is obtained by regressing the forecast variable on the in-sample combination forecasts and a constant:

$$r_{s+1} = \eta + \delta \hat{r}^C_{s+1} + \varepsilon_{s+1}, \tag{6}$$

implying that an out-of-sample iterated combination forecast $\hat{r}^{IC}_{t+1}$ made at time t takes the following form:

$$\hat{r}^{IC}_{t+1} = \hat{\eta}_t + \hat{\delta}_t \hat{r}^C_{t+1}. \tag{7}$$

We construct the *iterated mean combination* (IterMean), the *iterated median combination* (IterMedian) and the *iterated trimmed mean combination* (IterTrimmed).[2]

[2]Lin, Wu, and Zhou (2018) focus on the *iterated weighted average* (IterWeighted) forecast, where the weights in the first combination step are determined in proportion to the inverse of estimated residual variances up to time t. We find that results for the IterWeighted forecast are virtually indistinguishable from the IterMean forecast in our data sets. Furthermore, note that the IterMean forecast coincides with the *partial least squares* (PLS) forecast.

2.3 Penalized regression approaches

Penalized regression approaches extend the objective function of the linear model given by equation (1) by adding a penalty term to shrink the predictor coefficients toward zero. Shrinkage is used to mitigate overfitting and improve the model's out-of-sample performance.

Elastic net, LASSO and ridge regression

The *elastic net* (ENet) estimates of the coefficients $(\alpha, \beta_1, ..., \beta_N)$ are chosen to minimize the following equation:

$$\sum_{s=1}^{t-1} \left(r_{s+1} - \alpha - \sum_{i=1}^{N} \beta_i x_{i,s} \right)^2 + \lambda \left[\frac{1}{2}(1-\delta) \sum_{i=1}^{N} \beta_i^2 + \delta \sum_{i=1}^{N} |\beta_i| \right], \tag{8}$$

where $\lambda > 0$ is a regularization parameter controlling the extent of shrinkage and $0 \leq \delta \leq 1$ is a parameter that serves to determine the respective weight of the two penalty terms $\sum_{i=1}^{N} \beta_i^2$ and $\sum_{i=1}^{N} |\beta_i|$. The first component is the sum of squared residuals, implying that the ENet will produce the OLS estimates when $\lambda = 0$. For the special case that $\delta = 0$, the method is called ridge regression, while $\delta = 1$ gives the *least absolute shrinkage and selection operator* (LASSO). An important distinction is that the LASSO forces the coefficients on some predictors to be exactly equal to zero, meaning that it can be employed as a technique to select variables from a larger group of potential predictors. For values of δ between 0 and 1, one gets a blend of LASSO and ridge regression that performs both selection and shrinkage. In the weighted least squares variant that we consider, we first weight returns and predictors with the inverse of the estimated variance before applying the penalized estimation routines.

Out-of-sample performance depends on the choice of the regularization parameter λ. We follow Rapach and Zhou (2020) in choosing λ according to the *corrected Akaike information criterion* (AICC) and setting $\delta = 0.5$ in the ENet forecast.

Combination elastic net

Rapach and Zhou (2020) suggest that the mean combining method can be improved by selecting only the relevant univariate forecasts. For the selection, they estimate the following regressions via the ENet:

$$r_{s+1} = \eta + \sum_{i=1}^{N} \zeta_i \hat{r}_{i,s+1} + \varepsilon_{s+1} \tag{9}$$

To select forecasts for an out-of-sample study, the regressions are run recursively, over an expanding hold-out period $s = m, ..., t-1$, which we initially set to ten years. Following Rapach and Zhou (2020), we restrict all ζ_i to be non-negative, which rules out that a selected univariate forecast is negatively related to the equity premium.

Defining $\mathcal{J}_t \subseteq \{1, ..., N\}$ as the index set of selected univariate forecasts, we obtain the *combination Elastic Net* (Comb-ENet) forecast for $t+1$ as

$$\hat{r}_{t+1}^{C-ENet} = \begin{cases} \frac{1}{|\mathcal{J}_t|} \sum_{i \in \mathcal{J}_t} \hat{r}_{i,t+1} & |\mathcal{J}_t| \geq 1 \\ \bar{r}_t & \text{else,} \end{cases} \tag{10}$$

where $|\mathcal{J}_t|$ is the cardinality of $\mathcal{J}_t$ and $\hat{r}_{i,t+1}$ is a univariate forecast comprised in the selection.[3]

2.4 Dimension reduction techniques

Techniques to reduce the dimension of the predictor data set provide an alternative approach to decrease the number of parameters that have to be estimated and to avoid an overfitted model.

Principal component regression

Principal component regression involves the estimation of a predictive model where the first K principal components extracted from a set of N predictors are used as the regressors, with $K \ll N$. The principal components $P_{k,t}$ serve as input in a standard predictive regression model:

$$r_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k P_{k,t} + \varepsilon_t. \tag{11}$$

Since overfitting becomes a concern as the number of estimated parameters increases, the literature advises to keep K fairly small (Rapach and Zhou (2013)). We experiment with two different specifications. First, we implement a sparse model that only makes

[3]Note that equation (10) differs from the definition given in Rapach and Zhou (2020) as we have to make allowance for the fact that none of the forecasts might be selected by the ENet.

use of the first principal component (PrinComp(1)), in accordance with Rapach and Zhou (2013) or Dichtl et al. (2021). Adopting an alternative specification that is used in Neely et al. (2014), we carry out an implementation (PrinComp(Opt)) where K is selected recursively according to the adjusted R^2. As in Neely et al. (2014), we restrict K to a maximum value of three to encourage a parsimonious model.

The three-pass regression filter

Principal component regression ignores any information on how the predictors are related to the forecasting target (Gu, Kelly, and Xiu (2020)). In contrast, the *three-pass regression filter* (3PassFilter) of Kelly and Pruitt (2015) provides a mechanism to directly connect the dimension reduction with the forecasting objective. Forecasting returns at time t using the 3PassFilter involves the following steps:

1. Run time series regressions of each lagged predictor $x_{i,s-1}$ on the realized equity premia r_s and a constant; save the estimated slopes denoted by $\hat{\phi}_{i,t}$.
2. For each period in $s = 1, ..., t$, run cross-sectional regressions of the predictors $x_{i,s}$ on the previously obtained slope estimates $\hat{\phi}_{i,t}$ and a constant; save the estimated slope coefficients $\hat{F}_s$.
3. Run the standard predictive time series regression of equity premia r_s on $\hat{F}_{s-1}$ and a constant, for $s = 2, ..., t$ to obtain the equity premium forecast for $t + 1$.

By default, all regressions are estimated via OLS. We also compute an alternative forecast where we use WLS in the *time series* regressions.[4]

The modified three-pass regression filter

Biases that arise from p-hacking are a major concern in the context of financial forecasting. Rytchkov and Zhong (2020) consequently propose a modified version of the 3PassFilter procedure, dubbed 3PassFilterM here, which they argue to be more conservative in the face of p-hacking. In the implementation of 3PassFilterM, the original predictors are replaced with their studentized univariate return forecasts $\hat{r}_{s+1}$ in steps 1 and 2 of the 3PassFilter.

[4]In addition, we scale the predictors to have a mean of zero and unit variance. Note that in this case, the 3PassFilter would coincide with the PLS and thus the IterMean forecast if no constant is included in the first two regressions.

2.5 Model restrictions

As suggested by Campbell and Thompson (2008), we examine whether restrictions on the signs of the regression coefficients and return forecasts help to improve out-of-sample performance. We proceed as follows: for the combination forecasts, we set the slopes in the recursive univariate regressions to zero if they do not match expected signs. The latter are determined with in-sample regressions. We also require the univariate equity premium forecasts to be non-negative. These adjustments are implemented before combining.

In the case of the kitchen sink, ENet, LASSO, and ridge forecasts, we estimate all regressions subject to the coefficient constraints on each predictor. If a forecast turns out to be negative, we set it to zero. For the iterated combinations and the dimension reduction methods, we constrain the forecast to be non-negative.

3 The evaluation of forecasts

In this section, we describe the statistical and economical criteria that we use to evaluate forecast quality.

3.1 Statistical criteria

We compare the performance of the competing forecasts to a benchmark forecast. As is standard in the literature, this benchmark is the average $\bar{r}_t$ of the equity premium until time t, referred to as the historical average or prevailing mean forecast. Our main statistical measure is the out-of-sample R^2, which is defined as

$$R^2_{OS} = 1 - \frac{\sum_{t=m}^{T-1}(r_{t+1} - \hat{r}_{t+1})^2}{\sum_{t=m}^{T-1}(r_{t+1} - \bar{r}_t)^2}, \tag{12}$$

where m is the first period in which an out-of-sample forecast is made. The R^2_{OS} statistic can be interpreted as the reduction in the mean squared prediction error (MSPE) that a forecast $\hat{r}_{t+1}$ achieves relative to the historical average.

The econometric literature has developed methods for testing whether the reduction in MSPE achieved by a given predictive model compared to a reference forecast is statistically significant. For nested models, the MSPE-adjusted statistic proposed by Clark and West (2007) is routinely employed. One starts by defining

$$\widehat{test}_{t+1} = (r_{t+1} - \bar{r}_t) - \left[(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_t - \hat{r}_{t+1})^2\right]. \tag{13}$$

Next, $\widehat{test}_{t+1}$ is regressed on a constant and the t-statistic for the constant is used to obtain a p-value for a one-sided upper tail test with the standard normal distribution. This corresponds to a test of the null hypothesis that $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$.

3.2 Economic criteria

Statistical metrics based on MSPE are the most common criteria to gauge forecast quality. However, there is no one-to-one relationship between the MSPE and the economic benefits of using the forecast (cf. Rapach and Zhou (2013)). For this reason, we follow the literature and also calculate the gains in utility achieved by a mean-variance investor. We set the relative risk aversion parameter γ equal to three, as in Rapach, Strauss, and Zhou (2010).

Assume that in each time period, an investor allocates her portfolio between the stock market and the risk-free Treasury bill. The asset allocation can either be performed based on the historical average or the alternative model forecast. The investor will allot the following proportion of her portfolio in period $t+1$ to the stock market:

$$\omega_{p,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\mathbb{E}\left[r_{t+1}\right]}{\mathbb{E}\left[\sigma^2_{t+1}\right]}\right), \tag{14}$$

where $\omega_{p,t}$ denotes the weight of stocks in the portfolio and the expected value $\mathbb{E}$ of the equity premium in $t+1$ is either given by the historical average $\bar{r}_t$ or the alternative forecast $\hat{r}_{t+1}$. As in Rapach, Strauss, and Zhou (2010), we use a ten-year rolling window to estimate the variance σ^2_{t+1} of equity premia, and constrain the portfolio weights on the stock market to be in the interval between 0 and 1.5. This means that short selling is not allowed and that leverage is limited to a maximum of 50 %.[5] Having thus allotted a share of her portfolio in each time period t, the investor has realized the following average utility level $\hat{\nu}_p$:

$$\hat{\nu}_p = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p, \tag{15}$$

[5] In the asset allocation exercise, we use the simple instead of the log return. In this fashion, the return on the portfolio in $t+1$ can conveniently be calculated as the product of the portfolio weights and the stock return, plus the risk-free rate.

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are, respectively, the sample mean and variance of the return on a portfolio constructed either using the historical average or the alternative forecast.

In the literature, there is no widely used convention for dealing with transaction costs. Rapach, Strauss, and Zhou (2010) do not report utility gains with transactions costs. Other studies report results assuming proportional transaction costs of 25 basis points (Dichtl et al. (2021)) or 50 basis points (Li and Tsiakas (2017), Neely et al. (2014)). For the main results, we will assume proportional transaction costs of 20 basis points (bps).

The utility gain of an investor that uses the alternative instead of the benchmark forecast is then given as the annualized difference between the respective average utility levels. The utility gain has a very intuitive economic interpretation as the management fee that an investor would be willing to pay to have access to the additional information from the alternative forecast that is not contained in the historical average forecast.

4 Empirical Analysis

4.1 Data

The variable that we forecast is the US equity premium, measured as the difference between the return on the S&P 500 index (dividends included) and the risk-free return. The latter is taken to be the interest rate at which Treasury bills with a maturity of three months are sold on the secondary market. As is common in the literature, we use continuously compounded returns for the statistical evaluation of forecast errors, and simple returns for the computation of utility gains.[6]

For the main analysis, we use the 15 predictors used in Rapach, Strauss, and Zhou (2010) for the prediction of the quarterly equity premium. For the monthly frequency, we use the same 14-predictor subset as in Rapach and Zhou (2013). A detailed definition of the variables is provided in Appendix A.

The data are taken from the 2020 version of the data file maintained by Amit Goyal. The predictors and these data set are not only used in the seminal paper of Welch and Goyal (2008) but also in most subsequent predictability studies.

[6]Rapach, Strauss, and Zhou (2010) state explicitly that they use continuously compounded returns for the analysis of forecast errors, but do not mention that they use simple returns for the asset allocation study. Simple returns, however, are not only the theoretically correct input given the assumed utility function, they are also used for calculating the utility gains of combination forecasts in Rapach and Zhou (2013).

4.2 Forecast performance

Quarterly frequency

We first present results for the setup chosen in Rapach, Strauss, and Zhou (2010): The variable to be forecasted is the quarterly equity premium, the first out-of-sample forecasts is made in 1964:4 for 1965:1, and the predictor data start in 1947:1, when the last of the 15 used predictors becomes available. As mentioned in previous sessions, other choices such as the length of the hold-out period for DMSPE combination forecasts, or the risk aversion and allocation bounds assumed for the asset allocation exercise are chosen as in Rapach, Strauss, and Zhou (2010). The only difference is that we report the utility gains after taking proportional transaction costs of 20 basis points into account.

Table 1 shows the summary statistics for the statistical and economic performance of the forecast methods we study when no Campbell-Thompson restrictions are applied. We provide results for the entire out-of-sample period from 1965:1 to 2020:4 as well as for two equally-sized subperiods. Over the entire sample period, the combination methods studied in Rapach, Strauss, and Zhou (2010), which are indicated by the acronyms Mean, Median, Trimmed and DMSPE(θ=1) and D(θ=.9) lead to significantly greater statistical forecast accuracy as well as to positive utility gains. A look at the subperiods, however, shows that this advantage over the historical mean benchmark is driven by the first half of the out-of-sample period. From 1993:1 to 2020:4, each combination forecast studied by Rapach, Strauss, and Zhou (2010) shows a negative out-of-sample R^2 and a negative utility gain. Results do not depend on whether the regressions with which the individual predictions are generated are estimated with the standard least squares approach or with a weighted least squares approach as in Johnson (2018).

Variations of the combination approach such as iterated combination forecasts or combination elastic net forecasts as well as alternative approaches for dimension reduction and variable selection show the same pattern. From 1965 to 1992, several of these forecast methods lead to gains relative to the historical mean. In the second half of the out-of-sample period, however, none of the methods leads to forecasts that are significantly more accurate.

For each combination forecast method examined in Table 1, Figure 1 graphs the differences between the cumulative squared prediction error (CSPE) of the prevailing mean benchmark and the cumulative squared prediction error of the forecast. Figure 2 does the same for the other forecasts. If the line in a graph is positively sloped in some period, then a given method delivers more accurate predictions than the historical average in that period, while a negative slope indicates that the model was outperformed by the benchmark. An ideal forecast that always has a lower squared prediction error than the benchmark would therefore exhibit a consistently positive slope. When a slope ends above the black horizontal line, the overall CSPE difference is positive and the model predicted the equity premium more accurately than the benchmark.

With the help of the graphs, one can for example check whether the summary statistics presented in Table 1 for 1993:1 to 2020:4 are representative of the performance in more recent years, or to which extent they are affected by a few years. Overall, the graphs suggests that the summary statistics do not mislead about developments in some parts of that period. From the early 1990s on, the CSPE profiles are mostly negatively sloped or flat. The most recent years may be of special interest. From 2011 to 2020, for example, none of the methods shows a marked performance gain relative to the historical mean. Finally, it is interesting to note that none of the combination forecasts displays a remarkable performance during the financial crisis of 2007-2009 or the Corona crisis. Apparently, prior findings that predictability is particularly strong during recessions (e.g. Henkel, Martin, and Nardari (2011) and Rapach and Zhou (2013)) do not apply to more recent years.

Table 2 shows the summary statistics when restrictions on the sign of regression coefficients and the sign of the equity premium forecast are implemented as suggested by Campbell and Thompson (2008). Patterns and conclusions do not change. None of the out-of-sample R^2 in the second half of the sample period is significant on a level of 5%. The highest out-of-sample R^2 is achieved by the combination elastic net, with a significance of 10%. However, the R^2 value is quite low compared to the R^2 values reported for combination forecasts in Rapach, Strauss, and Zhou (2010), and the economic gain is just 0.01% per annum.

Monthly frequency

In the recent predictability literature, monthly prediction horizons are more common than the quarterly horizon employed by Rapach, Strauss, and Zhou (2010). Table 3 therefore shows the forecast performance for the monthly horizon, now again without Campbell-Thompson restrictions. As one of the predictors used before set is not available on a quarterly frequency, the number of predictors reduces from 15 to 14. As is evident from the table, conclusions do not change when we move to the monthly frequency: None of the forecasts methods leads to gains relative to the benchmark. Unreported results in which we apply the Campbell-Thompson restrictions show the same patterns.

4.3 Sensitivity analysis

For the presentation of our main results, we already conducted a lot of variations. Though our focus is on the methods employed in Rapach, Strauss, and Zhou (2010), we considered related forecasting approaches as well as the weighted-least squares approach suggested by Johnson (2018). To further demonstrate the robustness of the results, we consider the following additional variations:

Extending the estimation sample

For the monthly forecast horizon, we extend the start of the estimation sample to 1926:12, as in Rapach and Zhou (2013). Conclusions do not change. Out of all methods considered Table 3, the highest R^2 we obtain for 1993:01 to 2020:12 is 0.16, significant at the 10% level and associated with a negative utility gain of -0.41.

Predictor set of Li and Tsiakas (2017)

Several of the predictor variables are highly correlated or even linearly dependent. For their study of monthly return predictability, Li and Tsiakas (2017), who focus on Ridge, LASSO and elastic net regressions, exclude the dividend–price ratio, the payout ratio, and the long-term yield. When we re-do the monthly analysis with this reduced set of 11 predictors, the highest R^2 we obtain for 1993:01 to 2020:12 with the regularization methods studied in Li and Tsiakas (2017) is -3.38, associated with a utility loss of -1.85.

Predictor set of Rapach and Zhou (2020)

Rapach and Zhou (2020) do not only suggest the combination elastic net as a new forecasting method, they also use a different predictor set for their empirical analysis. In comprises nine financial and macroeconomic variables plus three technical indicators. When we use these variables instead of the 14 variables used in Table 3, the R^2 we obtain for the combination elastic net forecast suggested in Rapach and Zhou (2020) improves only slightly, from -0.16 to 0.05. The utility gain is still negative (-0.86).

The additional variations considered here therefore do not lead to a change in conclusions: Even though we consider a large number of different forecasting methods and several variations related to data frequency, estimation sample, and predictor sets, we fail to find significant predictability after 1992.

4.4 Why has predictive performance declined?

Rapach, Strauss, and Zhou (2010) offer two explanations for the success of combination forecasts they document in their sample: "(i) combining forecasts incorporates information from numerous economic variables while substantially reducing forecast volatility; (ii) combination forecasts are linked to the real economy." (Rapach, Strauss, and Zhou (2010), p. 821).

The first-mentioned feature of combination forecasts is still present in the 1993 to 2020 period. As in earlier years, combining is a way of aggregating information from many variables, and unreported results also confirm that combining shrinks the volatility of forecasts relative to the volatility of individual forecasts. Rapach and Zhou (2020) discuss that the benefits of shrinkage can start to decline once it gets too strong. A possible explanation for the decline in the performance of combination forecasts is therefore that the shrinkage moved beyond the optimal point. However, the solution that Rapach and Zhou (2020) offer for tuning shrinkage—the combination elastic net—does not lead to a clear improvement over the simple combination methods. Without the Campbell-Thompson restrictions (cf. Table 1), it leads to a smaller accuracy than combination forecasts studied in Rapach, Strauss, and Zhou (2010); with restrictions (cf. Table 2), the performance is better but still not significant at a 5% level, with a utility gain of only 0.01%. Though the shrinkage aspect may play a role in explaining the performance decline, it is not evident from the literature how this insight could be used to generate forecasts that are superior to the historical mean benchmark.

To investigate whether the links to the real economy that Rapach, Strauss, and Zhou (2010) document have changed over time, we study how useful equity premium forecasts are for the prediction of industrial production and real GDP growth. Industrial production data have the benefit of being available on a monthly frequency for a sufficiently long time period. For the monthly analysis of real GDP growth we estimate monthly growth rates using cubic interpolation. Data on GDP and industrial production are obtained from the FRED database.

As in Rapach, Strauss, and Zhou (2010), we use an autoregressive distributed lag model

$$y_{t+1} = \eta_j + \phi y_t + \beta_j x_{j,t} + \varepsilon_{t+1}. \tag{16}$$

where $y_t + 1$ denotes the growth rate of real GDP or the industrial production index from period t to $t+1$ and $x_{j,t}$ is an out-of-sample combination forecast. We estimate (16) recursively using ordinary least squares to generate out-of-sample forecasts of the two macroeconomic variables. In addition to monthly forecasts, we consider 12-month ahead forecasts, for which we replace y_{t+1} with $y_{t+12} = (1/12)(y_{t+1} + ... + y_{t+12})$. Finally, we determine the out-of-sample R^2 of the forecasts based on (16) relative to the historical means of the real GDP and industrial production growth rates.

Table 4 reports these R^2 values for the same periods that we studied previously. The time pattern in the predictability of industrial production growth is the same as in the one in the predictability of equity premia: Combination forecasts provide a superior accuracy from 1965 to 1992 but fail to deliver in the later period from 1993 to 2020. Forecasts of real GDP growth rates also show the tendency to provide smaller gains relative to the benchmark, even though some of the combination forecasts continue to contribute to significant forecast gains in the second half of the period.

We conclude that the weaker link between equity premium forecasts and the real economy is consistent with the declining gains from using these forecasts for equity premium prediction. Though the relationship is not necessarily causal, the finding indicates that there could be a fundamental change in the information content of combination forecasts.

5 Conclusion

In the present paper, we reappraise the ability of combination forecasts to predict the US equity premium in out-of-sample tests over the period from 1965 to 2020. For the first half of this period, we confirm the finding of Rapach, Strauss, and Zhou (2010): From 1965 to 1992, combination forecasts predict the equity premium significantly better than the historical average benchmark. For the more recent time period, from 1993 to 2020, however, combination forecasts do not provide benefits relative to the historical average. Modifying combination forecasts in ways suggested in other studies or using alternative approaches for dimension reduction and variable selection do not change the conclusion.

We also show that the lower quality of combination forecasts is accompanied by a weakening of their ability to predict the real economy. In addition, the degree of shrinkage may play a role for the success of combination forecasts. It is not obvious how these observations can lead to a revival of combination forecasts. The combination elastic net forecast that Rapach and Zhou (2020) offer as a solution to the problem of determining a suitable degree of shrinkage also fails to significantly outperform the historical mean from 1993 to 2020.

Given that the empirical performance of combination forecasts changed fairly quickly in the past, it may again improve in the future. However, there is also a plausible reason why predictive power could remain low: As shown by Rytchkov and Zhong (2020), collective data snooping by the academic community can introduce a data-snooping bias in combination forecasts. If their good performance in the 1965 to 1992 was spurious, it would not be surprising if combination forecasts failed to deliver in the future.

Though it is interesting to consider different explanations and scenarios for the performance of predictive models, it is not obvious how this can help investors and others choose a better forecast model. Consistent with this view, the literature on equity premium forecasts focuses on studying their empirical performance, rather than on pondering possible explanations for preferring one model over the other.

Looking at the data available to them, Rapach, Strauss, and Zhou (2010) concluded that combination forecasts help to predict the equity premium. Looking at the data that are available today and that we analyzed in the present paper, the conclusion that researchers or investors reach may well be different.

A Description of variables

The following 15 variables constitute the predictor set for the quarterly prediction horizon. For monthly predictions, we consider only the first 14 variables. The expected signs used in applying the Campbell-Thompson restrictions are given in parentheses.

- ◇ Log dividend-price ratio (dp): the difference between the log of a 12-month moving sum of dividends paid on the S&P 500 index, and the log of stock prices as measured by the S&P 500. (+)
- ◇ Log dividend yield (dy): the difference between the log of a 12-month moving sum of dividends paid on the S&P 500 index, and the log of the lagged S&P 500 price index. (+)
- ◇ Log price-earnings ratio (ep): the difference between the log of a 12-month moving sum of earnings on the S&P 500, and the log of the S&P 500. (+)
- ◇ Log dividend-payout ratio (de): the log of a 12-month moving sum of dividends minus the log of a 12-month moving sum of earnings. (+)
- ◇ Stock variance ($svar$): the sum of squared daily returns on the S&P 500. (+)
- ◇ Book-to-market ratio (bm): the ratio of book value to market value for the Dow Jones Industrial Average (DJIA). (+)
- ◇ Net equity expansion ($ntis$): a 12-month moving sum of net equity issues by stocks listed on the New York Stock Exchange (NYSE) divided by the total end-of-year market capitalization of NYSE stocks. (−)
- ◇ Treasury bill rate (tbl): the interest rate on a three-month Treasury bill in the secondary market. (−)
- ◇ Long-term yield (lty): the yield on long-term government bonds. (−)
- ◇ Long-term return (ltr): the rate of return on long-term government bonds. (+)
- ◇ Term spread (tms): the long-term yield minus the Treasury bill rate. (+)
- ◇ Default yield spread (dfy): the difference between BAA- and AAA-rated corporate bond yields. (+)
- ◇ Default return spread (dfr): the return on long-term corporate bonds minus the return on long-term government bonds. (+)

◇ Inflation (*infl*): the growth in the Consumer Price Index (CPI, all urban consumers). Inflation is lagged by an additional period to account for the fact that CPI data is only released in the following quarter. (−)

◇ Investment-to-capital ratio: Ratio of aggregate (private nonresidential fixed) investment to aggregate capital.(+)

The data for the construction of the equity premium and the predictors are taken from the website of Amit Goyal.[7] For a more detailed description of the variables, we refer to Welch and Goyal (2008).

[7] The data are available from http://www.hec.unil.ch/agoyal/.

References

Campbell, John and Samuel B. Thompson (2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" In: *Review of Financial Studies* 21.4, pp. 1509–1531.

Clark, Todd E. and Kenneth D. West (2007). "Approximately normal tests for equal predictive accuracy in nested models". In: *Journal of Econometrics* 138.1, pp. 291 – 311.

Dichtl, Hubert, Wolfgang Drobetz, Andreas Neuhierl, and Viktoria-Sophie Wendt (2021). "Data snooping in equity premium prediction". In: *International Journal of Forecasting* 37.1, pp. 72 –94.

Farmer, Leland, Lawrence Schmidt, and Allan Timmermann (2019). "Pockets of predictability". In: *Available at SSRN 3152386*.

Goyal, Amit, Ivo Welch, and Athanasse Zafirov (2021). *A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II*. Tech. rep. Swiss Finance Institute.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2020). "Empirical Asset Pricing via Machine Learning". In: *Review of Financial Studies* 33.5, pp. 2223–2273.

Harvey, David I, Stephen J Leybourne, Robert Sollis, and AM Robert Taylor (2021). "Real-time detection of regimes of predictability in the US equity premium". In: *Journal of Applied Econometrics* 36.1, pp. 45–70.

Hendry, David F and Michael P Clements (2004). "Pooling of forecasts". In: *Econometrics Journal* 7.1.

Henkel, Sam James, J. Spencer Martin, and Federico Nardari (2011). "Time-varying short-horizon predictability". In: *Journal of Financial Economics* 99.3, pp. 560–580.

Johnson, Travis L. (2018). "A Fresh Look at Return Predictability Using a More Efficient Estimator". In: *Review of Asset Pricing Studies* 9.1, pp. 1–46.

Kelly, Bryan and Seth Pruitt (2015). "The three-pass regression filter: A new approach to forecasting using many predictors". In: *Journal of Econometrics* 186.2, pp. 294 – 316.

Lettau, Martin and Stijn Van Nieuwerburgh (2008). "Reconciling the return predictability evidence". In: *The Review of Financial Studies* 21.4, pp. 1607–1652.

Li, Jiahan and Ilias Tsiakas (2017). "Equity premium prediction: The role of economic and statistical constraints". In: *Journal of Financial Markets* 36.C, pp. 56–75.

Lin, Hai, Chunchi Wu, and Guofu Zhou (2018). "Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach". In: *Management Science* 64.9, pp. 4218–4238.

Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou (2014). "Forecasting the Equity Risk Premium: The Role of Technical Indicators". In: *Management Science* 60.7, pp. 1772–1791.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Pastor, Lubos and Robert F Stambaugh (2001). "The equity premium and structural breaks". In: *The Journal of Finance* 56.4, pp. 1207–1239.

Rapach, David E., Jack Strauss, and Guofu Zhou (2010). "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy". In: *Review of Financial Studies* 23.2, pp. 821–862.

Rapach, David E. and Guofu Zhou (2013). "Forecasting Stock Returns". In: *Handbook of Economic Forecasting*. Ed. by Graham Elliott, Clive Granger, and Allan Timmermann. Vol. 2A. Amsterdam: Elsevier, pp. 328–383.

– (2020). "Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods". In: *Machine Learning for Asset Management*. Ed. by Emmanuel Jurczenko. Hoboken: John Wiley & Sons, Ltd., pp. 1–33.

Rytchkov, Oleg and Xun Zhong (2020). "Information Aggregation and P-Hacking". In: *Management Science* 66.4, pp. 1605–1626.

Welch, Ivo and Amit Goyal (2008). "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". In: *Review of Financial Studies* 21.4, pp. 1455–1508.

Table 1: Out-of-sample performance of quarterly equity premium forecasts without Campbell-Thompson restrictions and estimation sample start in 1947

	1965:1-2020:4				1965:1-1992:4				1993:1-2020:4			
	LS		WLS		LS		WLS		LS		WLS	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Method	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ
KitchenSink	−23.72*	0.64	−20.88**	1.11	−13.42***	5.16	−13.68***	4.58	−34.58	−3.79	−28.49	−2.32
Combination Forecasts												
Mean	2.42***	1.60	2.80***	1.91	5.07***	3.49	5.52***	3.98	−0.38	−0.28	−0.06	−0.14
Median	1.70***	0.51	2.31***	0.97	4.00***	1.74	4.73***	2.25	−0.72	−0.73	−0.24	−0.31
Trimmed	2.35***	1.45	2.72***	1.76	4.91***	3.18	5.31***	3.62	−0.34	−0.28	−0.02	−0.09
IterMean	−1.49***	3.29	0.19***	3.61	7.99***	7.54	8.46***	7.61	−11.49	−0.85	−8.55	−0.30
IterMedian	−0.68**	1.63	1.87***	3.60	8.62***	6.44	10.18***	7.31	−10.48	−3.09	−6.90	−0.06
IterTrimmed	0.79***	3.58	2.21***	4.12	11.51***	8.03	11.81***	8.23	−10.52	−0.76	−7.92	0.09
DMSPE(θ=1)	2.39***	1.54	2.79***	1.87	5.04***	3.40	5.51***	3.91	−0.41	−0.31	−0.09	−0.17
DMSPE(θ=.9)	2.36***	1.47	2.76***	1.79	4.92***	3.15	5.41***	3.68	−0.35	−0.21	−0.03	−0.09
DMSPE(θ=.5)	2.24***	1.41	2.65***	1.75	4.58***	2.79	5.06***	3.32	−0.23	0.04	0.10	0.18
Comb-ENet	0.61	1.40	0.77	1.46	2.35	4.52	2.77*	4.24	−1.23	−1.72	−1.35	−1.33
Other Methods												
ENet	−17.16	0.46	−12.31**	0.85	−8.95***	5.46	−9.15***	5.36	−25.83	−4.42	−15.66	−3.57
LASSO	−17.75	0.58	−12.06**	0.75	−10.22***	5.19	−8.61***	5.78	−25.70	−3.93	−15.71	−4.17
Ridge	−19.85**	0.74	−16.78**	1.46	−8.56***	5.31	−8.52***	5.01	−31.78	−3.72	−25.49	−2.06
PrinComp(1)	2.94***	2.92	3.12***	3.02	5.84***	5.23	6.07***	5.47	−0.11	0.62	0.01	0.60
PrinComp(Opt)	−2.81**	2.50	−1.86***	2.65	6.02***	8.17	5.64***	7.58	−12.13	−3.00	−9.77	−2.17
3PassFilter	0.52***	2.66	1.64***	2.97	8.07***	6.20	8.40***	6.21	−7.46	−0.84	−5.50	−0.25
3PassFilterM	−7.18	−1.16	−8.28	−1.27	−13.57	−1.47	−15.46	−1.09	−0.45	−0.80	−0.70	−1.41

Notes: The table reports results for forecasts of the quarterly US equity premium based on 15 individual predictors. In-sample estimation begins in 1947:1. Predictive regression models are either estimated with a least squares (LS) or weighted least squares (WLS) approach. R^2_{OS} denotes the reduction (in) of mean squared prediction error (MSPE) achieved by a forecast based on a method in column (1) relative to the trailing historical average. Statistical significance of the R^2_{OS} statistic is assessed using the Clark and West (2007) MSPE-adjusted statistic. *,**, and *** indicate significance at the 10 %, 5 % and 1 % levels, respectively. Δ is the utility gain (in % per annum) of a mean-variance investor with relative risk aversion coefficient $\gamma = 3$ who used a predictive regression forecast based on a method in column (1) instead of the historical average. Utility gains are reported net of a proportional transaction cost of 20 basis points.

Table 2: Out-of-sample performance of quarterly equity premium forecasts with Campbell-Thompson restrictions and estimation sample start in 1947

	1965:1-2020:4				1965:1-1992:4				1993:1-2020:4			
	LS		WLS		LS		WLS		LS		WLS	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Method	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ
KitchenSink	−7.79**	1.50	−8.56**	2.43	−1.93***	6.91	−2.20***	7.11	−13.97	−3.80	−15.26	−2.19
Combination Forecasts												
Mean	2.26***	1.60	2.47***	1.91	4.26***	3.49	4.60***	3.98	0.14	−0.28	0.22	−0.14
Median	1.91***	0.51	2.32***	0.97	4.18***	1.74	4.73***	2.25	−0.49	−0.73	−0.22	−0.31
Trimmed	2.29***	1.45	2.48***	1.76	4.44***	3.18	4.72***	3.62	0.02	−0.28	0.12	−0.09
IterMean	3.84***	3.29	4.10***	3.61	8.03***	7.54	8.01***	7.61	−0.58	−0.85	−0.03	−0.30
IterMedian	2.14***	1.63	3.43***	3.60	9.50***	6.44	10.70***	7.31	−5.63	−3.09	−4.23	−0.06
IterTrimmed	5.68***	3.58	6.22***	4.12	11.81***	8.03	11.84***	8.23	−0.79	−0.76	0.29	0.09
DMSEP(θ=1)	2.23***	0.89	2.45***	1.04	4.23***	2.02	4.58***	2.37	0.12	−0.24	0.20	−0.28
DMSEP(θ=.9)	2.21***	0.87	2.44***	1.00	4.16***	1.90	4.52***	2.22	0.15	−0.17	0.24	−0.21
DMSEP(θ=.5)	2.15***	0.89	2.40***	1.04	3.97***	1.73	4.35***	2.04	0.24	0.05	0.34	0.03
Comb-ENet	1.62**	1.87	1.53**	1.49	2.03*	3.72	2.25**	3.74	1.20*	0.01	0.77	−0.76
Other Methods												
ENet	−8.87*	0.54	−8.51**	1.32	−4.25***	5.50	−2.66***	6.01	−13.74	−4.30	−14.68	−3.29
LASSO	−7.61**	0.66	−8.58**	1.02	−5.81***	5.22	−3.95***	6.15	−9.51	−3.81	−13.47	−3.99
Ridge	−7.95**	1.29	−8.42**	2.04	−2.82***	6.52	−2.97***	6.53	−13.35	−3.83	−14.17	−2.39
PrincComp(1)	2.94***	2.92	2.96***	3.02	5.57***	5.23	5.66***	5.47	0.17	0.62	0.11	0.60
PrinComp(Opt)	1.39**	2.50	−4.11*	2.65	7.00***	8.17	3.28**	7.58	−4.54	−3.00	−11.91	−2.17
3PassFilter	3.74***	2.66	4.31***	2.97	7.80***	6.20	7.68***	6.21	−0.54	−0.84	0.75	−0.25
3PassFilter(M)	−2.92	−1.16	−2.79	−1.27	−3.88	−1.47	−3.46	−1.09	−1.92	−0.80	−2.08	−1.41

Notes: The table reports results for forecasts of the quarterly US equity premium based on 15 individual predictors. In-sample estimation begins in 1947:1. Predictive regression models are either estimated with a least squares (LS) or weighted least squares (WLS) approach. R^2_{OS} denotes the reduction (in) of mean squared prediction error (MSPE) achieved by a forecast based on a method in column (1) relative to the trailing historical average. Statistical significance of the R^2_{OS} statistic is assessed using the Clark and West (2007) MSPE-adjusted statistic. *,**, and *** indicate significance at the 10 %, 5 % and 1 % levels, respectively. Δ is the utility gain (in % per annum) of a mean-variance investor with relative risk aversion coefficient $\gamma = 3$ who used a predictive regression forecast based on a method in column (1) instead of the historical average. Utility gains are reported net of a proportional transaction cost of 20 basis points.

Table 3: Out-of-sample performance of monthly equity premium forecasts without Campbell-Thompson restrictions and estimation sample start in 1947

	1965:01-2020:12				1965:01-1992:12				1993:01-2020:12			
	LS		WLS		LS		WLS		LS		WLS	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Method	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ	R^2_{OS}	Δ
KitchenSink	−4.48**	3.17	−4.07**	3.06	3.74***	8.04	2.20***	7.67	−13.51	−1.67	−10.95	−1.53
Combination Forecasts												
Mean	0.87***	0.86	0.97***	1.13	1.94***	2.44	1.99***	2.74	−0.30	−0.72	−0.15	−0.48
Median	0.62***	0.20	0.64***	0.09	1.34***	1.05	1.30***	0.72	−0.17	−0.65	−0.09	−0.55
Trimmed	0.82***	0.77	0.99***	1.06	1.78***	2.12	1.95***	2.52	−0.23	−0.57	−0.07	−0.40
IterMean	−1.56**	3.14	−1.22**	2.99	4.30***	6.84	3.23***	7.31	−7.99	−0.53	−6.09	−1.30
IterMedian	−0.21**	0.31	−0.91*	0.13	2.71***	2.87	1.44***	1.87	−3.40	−2.23	−3.48	−1.61
IterTrimmed	−0.31**	2.11	0.28***	2.26	3.98***	5.81	4.13***	6.10	−5.02	−1.57	−3.94	−1.55
DMSPE(θ=1)	0.87***	0.84	0.97***	1.12	1.93***	2.41	1.99***	2.72	−0.31	−0.72	−0.15	−0.49
DMSPE(θ=.9)	0.86***	0.82	0.97***	1.09	1.91***	2.29	1.96***	2.58	−0.28	−0.64	−0.12	−0.40
DMSPE(θ=.5)	0.82***	0.74	0.92***	1.00	1.83***	2.10	1.88***	2.39	−0.29	−0.62	−0.14	−0.39
Comb-ENet	0.94**	1.97	1.27***	2.41	1.94***	4.72	2.76***	5.46	−0.16	−0.79	−0.36	−0.65
Other Methods												
ENet	−2.33**	−2.15	−3.92*	0.54	3.30***	−3.23	−0.92***	2.65	−8.51	−1.04	−7.22	−1.56
LASSO	−2.22**	−1.69	−3.50*	0.46	3.02***	−2.65	−0.27***	2.53	−7.97	−0.72	−7.04	−1.62
Ridge	−3.49**	−1.71	−4.16*	0.60	2.07***	−2.69	−0.78***	1.71	−9.58	−0.71	−7.86	−0.50
PrinComp(1)	0.68***	1.11	0.82***	1.26	1.52***	2.39	1.58***	2.43	−0.26	−0.18	−0.01	0.10
PrinComp(Opt)	−0.48***	2.33	−0.55***	2.98	1.08***	5.87	1.10***	6.10	−2.20	−1.18	−2.36	−0.12
3PassFilter	−1.07**	2.97	0.30***	2.33	3.71***	6.44	4.06***	6.60	−6.32	−0.47	−3.83	−1.93
3PassFilterM	−1.95	−1.97	−1.87	−1.97	−1.77	−1.48	−1.98	−1.70	−2.15	−2.46	−1.76	−2.24

Notes: The table reports results for forecasts of the monthly US equity premium based on 15 individual predictors. In-sample estimation begins in 1947:01. Predictive regression models are either estimated with a least squares (LS) or weighted least squares (WLS) approach. R^2_{OS} denotes the reduction (in) of mean squared prediction error (MSPE) achieved by a forecast based on a method in column (1) relative to the trailing historical average. Statistical significance of the R^2_{OS} statistic is assessed using the Clark and West (2007) MSPE-adjusted statistic. *,**, and *** indicate significance at the 10 %, 5 % and 1 % levels, respectively. Δ is the utility gain (in % per annum) of a mean-variance investor with relative risk aversion coefficient $\gamma = 3$ who used a predictive regression forecast based on a method in column (1) instead of the historical average. Utility gains are reported net of a proportional transaction cost of 20 basis points.

Table 4: How well do combination forecasts of the equity premium predict macroeconomic growth rates?

	One-month predictions			12-month predictions		
	1965-2020	1965-1992	1993-2020	1965-2020	1965-1992	1993-2020
Panel A: Prediction of industrial production growth						
Mean	3.85***	7.10***	2.09*	7.72***	14.30***	0.44
Median	1.84***	6.40***	−0.63	2.27**	4.14**	0.17
Trimmed	2.76***	7.01***	0.46*	5.58**	11.24**	−0.72
IterMean	7.81***	2.08***	10.91	−7.62***	1.33***	−13.30
IterMedian	−1.75***	7.50***	−6.76	5.63***	14.94**	−3.66*
IterTrimmed	3.11***	7.28***	0.85	−5.81**	4.25***	−14.23
DMSPE(θ=1)	1.60**	3.42***	0.61	1.43	7.35*	−5.37
DMSPE(θ=.9)	1.67**	3.47***	0.69	4.57**	9.50**	−1.07
DMSPE(θ=.5)	1.70**	3.55***	0.70	3.06*	4.55*	1.35
Comb-ENet	5.69***	7.37***	4.78*	15.86***	28.18***	2.50*
Panel B: Prediction of real GDP growth						
Mean	10.16***	15.17***	7.74***	10.04***	16.98***	1.37**
Median	5.64***	10.42***	3.33***	3.21***	5.83***	0.19
Trimmed	7.95***	14.53***	4.77***	8.12***	13.29***	1.59*
IterMean	22.89***	26.12***	21.33***	−2.05***	1.41***	−0.86**
IterMedian	13.88***	26.96***	7.56***	−7.41***	−13.03***	5.81**
IterTrimmed	15.66***	25.64***	10.84***	3.51***	6.70***	2.83**
DMSPE(θ=1)	9.53***	15.08***	6.85***	10.11***	17.06***	1.41**
DMSPE(θ=.9)	9.07***	13.41***	6.97***	10.54***	16.29***	3.56**
DMSPE(θ=.5)	9.91***	14.48***	7.70***	10.25***	15.19***	4.22**
Comb-ENet	14.33***	22.07***	10.60***	13.55***	25.19***	−0.91

Notes: The table reports results for out-of-sample forecasts of US industrial production (IP) growth and US real GDP growth. Forecasts are made with an autoregressive distributed lag model that includes a combination forecast of the equity premium as a predictor. In-sample estimation begins in 1947:01. R^2_{OS} denotes the reduction in mean squared prediction error (MSPE) achieved by a forecast based on a method in column (1) relative to the historical average forecast over the respective time period. Statistical significance of the R^2_{OS} statistic is assessed using the Clark and West (2007) MSPE-adjusted statistic that corresponds to a one-sided test of the null hypothesis that the MSPE of the historical average benchmark forecast is less than or equal to the competing forecast MSPE against the alternative that the historical average MSPE is greater than the competing forecast MSPE. *,**, and *** indicate significance at the 10 %, 5 % and 1 % levels, respectively. For h = 12, we use Newey and West (1987) standard errors with lag 12.

Figure 1: Cumulative squared error (CSPE) differences for quarterly combination forecasts of the equity premium (no Campbell-Thompson restriction)

Notes: This figure depicts differences in cumulative squared prediction errors (CSPE) for quarterly US equity premium forecasts based on various combining methods over the out-of-sample period of 1965:1-2020:4. The estimation period begins in 1947:1. The black (red) lines in each panel delineate the cumulative squared prediction error of the historical average forecast minus the cumulative squared prediction error of the combination forecast where the individual predictive regression models were estimated using OLS (WLS). Recessions as defined by the National Bureau of Economic Research (NBER) are signalled by vertical grey bars.

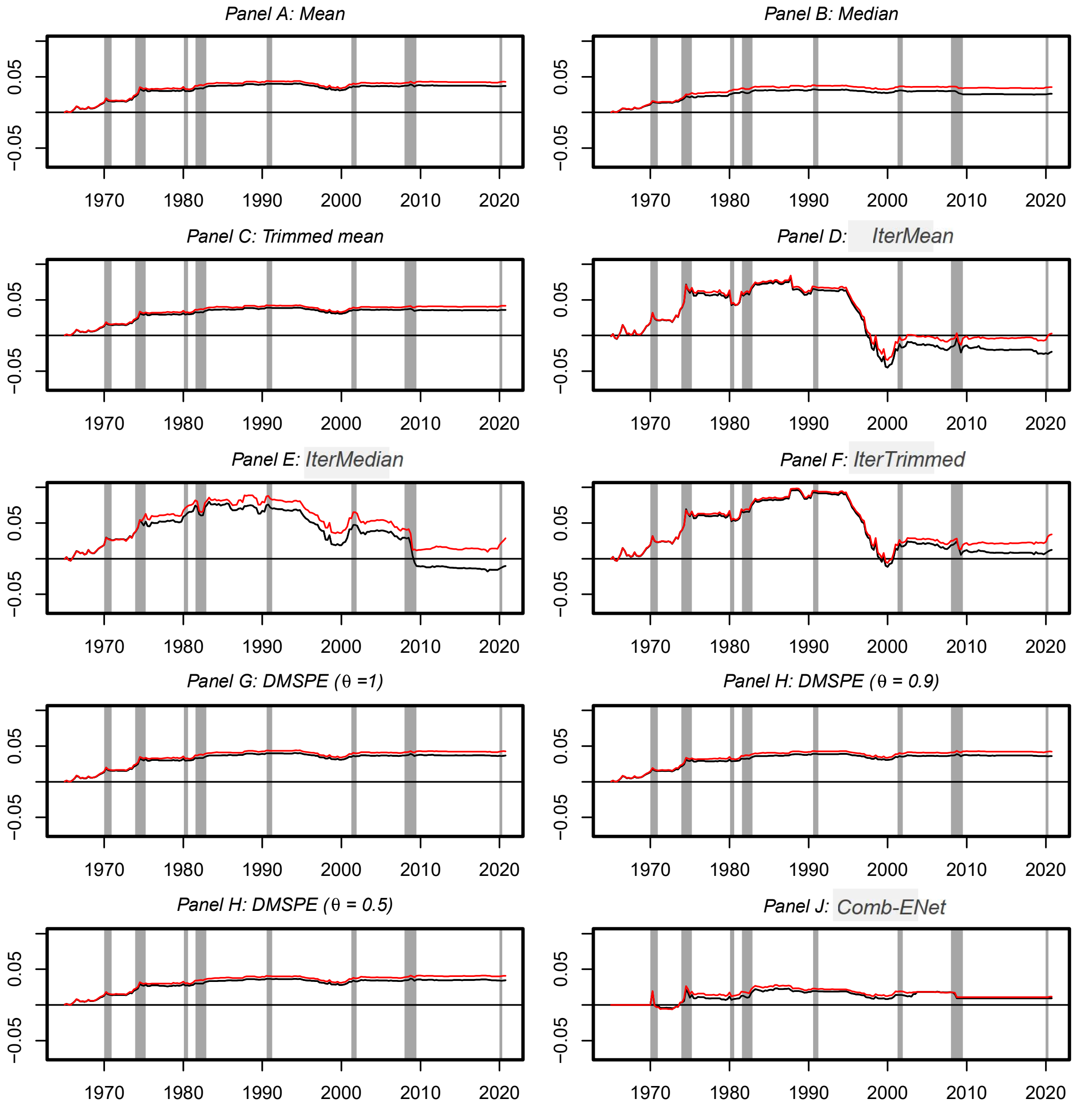

Figure 2: Cumulative squared error (CSPE) differences for alternative quarterly forecasts of the equity premium (no Campbell-Thompson restrictions)

Notes: This figure depicts differences in cumulative squared prediction errors (CSPE) for quarterly US equity premium forecasts based on various methods based on a multiple OLS kitchen sink regression, penalized regression and dimension reduction over the out-of-sample period of 1965:1-2020:4. The estimation period begins in 1947:1. The black (red) lines in each panel delineate the cumulative squared prediction error of the historical average forecast minus the cumulative squared prediction error of the competing forecast where the individual predictive regression models were estimated using least squares or weighted least squares. Recessions as defined by the National Bureau of Economic Research are signalled by vertical grey bars. Note the differences in scale for the kitchen sink, ENet, LASSO, Ridge and 3PassFilterM forecasts.