# Forecasting Stock Returns with Model Uncertainty and Parameter Instability*

Hongwei Zhang[1,2], Qiang He[1], Ben Jacobsen[2], and Fuwei Jiang[1]

[1]Central University of Finance and Economics – School of Finance

[2]Tilburg University – TIAS School for Business and Society

Asian Finance Association (AsianFA) 2018 Conference

First Draft: December 16 2017

This Draft: September 16 2019

*We acknowledge the very helpful comments and suggestions from David Rapach, Jan R. Magnus, Frans de Roon, Guofu Zhou, and seminar participants in Tilburg University, Central University of Finance and Economics, Tsinghua university, AsianFA 2018 Meeting. This article is supported by the National Natural Science Foundation of China (No. 71602198, 71572052), Beijing Natural Science Foundation (No. 9174045), and the Program for Innovation Research in Central University of Finance and Economics. Send correspondence to Hongwei Zhang, TIAS School for Business and Society, Tilburg University; e-mail: zhanghongwei@tsinghua.org.cn.

# Forecasting Stock Returns with Model Uncertainty and Parameter Instability

**Abstract**

We compare several representative sophisticated models of forecasting stock returns that accommodate model uncertainty, based on macroeconomic variables. When estimated traditionally, our results confirm that the simple combination of individual predictors is superior. However, sophisticated models improve dramatically once we combine them with the historical average and take parameter instability into account. Equal-weighted combination of the standard multivariate predictive regression with the historical average, for example, achieves a statistically significant monthly out-of-sample $R^2_{OS}$ of 1.10% and annual utility gains of 2.34%. We obtain similar gains for predicting factor portfolios and macro economic conditions.

**JEL classifications:** G17, G12, G02, C58

**Keywords:** Equity premia predictability, Forecast combination, Parameter instability

# 1 Introduction

Out-of-sample predictability of equity premia is at the core of financial economics (see, Rapach and Zhou, 2013, for a recent survey). As pointed out by Pesaran and Timmermann (1995) the main challenge results from model uncertainty and parameter instability.[1] To deal with model uncertainty, researchers in the forecasting literature have developed many sophisticated variable selection and model averaging techniques (e.g., the surveys by Hoeting et al., 1999; Timmermann, 2006; Claeskens and Jansen, 2008). However, most of these sophisticated models usually rely on a stationary environment, without structural breaks, as pointed out by Inoue et al. (2008). This is an important concern when applying these techniques to stock return forecasting based on macroeconomic variables, where parameters may be not stable and structural breaks are frequent (Stock and Watson, 1996).

By introducing a simple combining method, we compare different sophisticated models that deal with model uncertainty in the context of stock return forecasting. Inspired by Lin et al. (2017)'s iterated combination method for forecasting corporate bond returns, our approach also uses the historical average forecast as the shrinkage target but is a combination of the historical average with a sophisticated model. However, our approach differs in two ways. Firstly, rather than using the analytical optimal combination weights that are hard to estimate accurately in practice (Timmermann, 2006; Rapach and Zhou, 2013), our approach uses equal weights or performance-based weights.[2] In this sense, our approach is a natural extension of Rapach et al. (2010)'s simple (equal-weighted or performance-based) combining method for individual predictors. Secondly, to accommodate parameter instability, we estimate the sophisticated models in our approach using the average windows (AveW) forecasting method proposed by Pesaran and Timmermann (2007),

[1] Model uncertainty recognizes that a forecaster knows neither the best model specification nor its corresponding parameter values. Parameter instability means that the parameters in even the best model can vary over time because of structural breaks.

[2] The performance-based combination weights are computed based on the forecasting performance of individual models over a holdout out-of-sample period. Specifically, we use the discount mean squared forecast error (MSFE) of individual models to compute their combination weights.

instead of the traditional recursively expanding windows scheme. The AveW method averages the same models but computes these over different estimation windows to improve forecast accuracy in the presence of structural breaks. And while this approach has proved to be useful (Pesaran and Pick, 2011), we are not aware of an application to forecasting stock returns.

To test the effectiveness of our method, we consider several representative sophisticated models and find that it improves stock return forecasts of these models substantially. Specifically, we consider the popular BMA, weighted-average least squares (WALS) introduced by Magnus et al. (2010), Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO proposed by Tibshirani (1996) and Elastic net introduced by Zou and Hastie (2005). For comparison, we also consider the standard multivariate predictive regression known as the Kitchen sink model, and the simple combining methods for individual predictors from Rapach et al. (2010). We compare these models in the usual way in terms of the MSFE and economic gains using returns on the S&P 500 index. To make our results comparable with the literature, following Rapach et al. (2010), we use updated monthly data from Welch and Goyal (2008) over the period from 1926:12 to 2016:12. This data set also includes the usual suspects, such as the dividend-price ratio, the dividend yield, the Treasury bill rate, inflation, and the term spread.

When estimated traditionally, without taking parameter instability into account and using recursively expanding windows, none of these sophisticated models we investigate deliver a significantly positive out-of-sample $R^2_{OS}$. We can basically confirm that the simple combination of individual predictors is superior. When we incorporate parameter instability by using the average windows forecasting method, all sophisticated models we investigate reducing their MSFE modestly. In terms of the MSFE, however, all these models except LASSO and Elastic net still cannot outperform the simple combination of individual predictors. The relatively good performance of LASSO and Elastic net may result from their built-in shrinkage technique.

All these models can work if they are used in a shrinkage fashion as in Lin et al. (2017) and

Rapach et al. (2010), especially when they are estimated using the average windows forecasting method as in our simple combination scheme. Under our scheme, all these sophisticated models (except BMA) can outperform the simple combination of individual predictors in terms of both the MSFE and utility gains. For example, under our combination scheme using equal weights, WALS (Kitchen sink) achieves a $R^2_{OS}$ value of 1.22% (1.10%) and a utility gain of 2.46% (2.34%), both of which are much higher than that of the simple combination of individual predictors.

Contrary to the finding of Rapach et al. (2010), our results indicate that there is a strong predictability of stock returns in both NBER-dated business-cycle recessions (bad times) and expansions (good times). However, further investigation shows that forecasting gains of our combination scheme tend to be concentrated in periods of extreme stock market fluctuations, especially extreme market downturns. Our investigation confirms that the behavior of forecasts of our combination scheme agrees with the view of Fama and French (1989)and Campbell and Cochrane (1999); Cochrane (2007) that heightened risk aversion during economic downturns requires a higher risk premium.

With the same set of economic variables and using our combination scheme, we also find that these sophisticated models we investigate can significantly forecast the returns of the well-known 10 size characteristics portfolios, and many well-known macroeconomic variables, such as the Chicago Fed National Activity Index, Smoothed U.S. Recession Probabilities, the Output Gap and the Civilian Unemployment Rate.

With factors extracted from a macroeconomic database of 134 monthly U.S. indicators (Mccracken and Ng, 2015), instead of the above 12 macroeconomic variables from Welch and Goyal (2008), we find that our combination scheme can also substantially improve all these forecast models we investigate, implying the robustness of our results.

Intuitively, our combination scheme per se is a special kind of forecast combination, which can be viewed as a diversification strategy that improves forecasting performance in the same manner that asset diversification across a risk-free asset and a risky asset improves portfolio

performance, just as pointed out by Dunis et al. (2001); Timmermann (2006); Rapach and Zhou (2013) among others. The benefits of our scheme stem from diversification across models with different risk-return characteristics, where the sophisticated models estimated using the average windows forecast method serve as risky assets and the historical average forecast serves as a risky-free asset. By carefully choosing a shrinkage target (e.g. the historical average), our combination scheme helps sophisticated models suffering from overfitting to achieve a better bias-variance trade-off, resulting in a lower MSFEs. A forecast encompassing test further shows that the simple combinations of the historical average with different sophisticated models contain incremental forecasting information beyond the simple combination of individual predictors, and can significantly encompass the historical average forecast.

The first contribution of this paper is that we introduce a simple combination scheme which extends sophisticated models for model uncertainty, parameter instability and shrinkage simultaneously. Our combination scheme can be viewed as an extension of the simple combining method for individual predictors, but contains incremental forecasting information beyond it. The second contribution is that, based on our simple combination scheme, we apply the average windows forecasting method to stock return forecasting and confirm its effectiveness and robustness in handling parameter instability. The third contribution is that, we comprehensively compare several state-of-the-art sophisticated models, including the new WALS, JMA and the famous LASSO and Elastic net, in the context of stock return forecasting. We present empirical evidence regarding the usefulness of these sophisticated models for forecasting stock returns.

## 2 Econometric Methodology

### *2.1 Forecast Models*

We start with the standard multivariate regression model, called Kitchen sink, for predicting stock returns, which can be expressed as

$$r_{t+1} = X_t\beta + e_{t+1}, \quad t = 0, 1, ..., T-1 \tag{1}$$

where $r_{t+1}$ is the stock return in excess of the risk-free rate from the end of period $t$ to the end of period $t+1$, $X_t$ is a $1 \times (N+1)$ vector of predictors available at the end of $t$ that always includes a constant term as the first variable, the $(N+1) \times 1$ vector $\beta$ is the corresponding parameter vector, $e_{t+1}$ is a zero-mean disturbance term, $N$ is the number of predictors and $T$ is the sample size.

The standard model is simple but plagued by model uncertainty and parameter instability (e.g., Stock and Watson, 2006; Welch and Goyal, 2008). There are several ways to extend the standard model. We consider six representative variable selection or model averaging methods[3], as follows

(1) **Least Absolute Shrinkage and Selection Operator (LASSO)**: a popular variable selection and regularization method for performing shrinkage in regressions, introduced by Tibshirani (1996). It can improve the prediction accuracy and interpretability of regression models by altering the model fitting process to select only a subset of the provided variables for use in the final model rather than using all of them.

(2) **Elastic Net**: another variable selection and regularization method introduced by Zou and Hastie (2005), which overcomes some limitations of the LASSO algorithm by adding an $\iota_2$ penalty function. While enjoying a similar parsimonious representation, the Elastic net often

[3]Model averaging, also called forecast combination, when we concentrate on prediction rather estimation, is a popular method for dealing with model uncertainty. The idea underlying model averaging is that, instead of selecting the true or best possible model, we construct a combination forecast (or estimator) based on a weighted average over all candidate models. The weighted average forecast then incorporates model uncertainty in a natural way.

outperforms the LASSO.

(3) **Bayesian Model Averaging (BMA)**: a popular model averaging method for dealing with model uncertainty. The idea underlying BMA is that, instead of selecting the true or best possible model, we construct a combination forecast based on a weighted average over all candidate models, where the weights of candidate models are determined by their posterior probabilities.

(4) **Mallows Model Averaging (MMA)**: a frequentist model averaging technique, developed by Hansen (2007) and Wan et al. (2010), in which the model weights are selected by minimizing a Mallows criterion. Hansen (2008) shows that the Mallows criterion is an asymptotically unbiased estimator of both the in-sample mean squared error and the out-of-sample one-step-ahead MSFE. Unlike Bayesian methods such as BMA, frequentist model averaging methods such as MMA require no prior probabilities of models and the corresponding model weights are fully determined by the data.

(5) **Jackknife Model Averaging (JMA)**: another frequentist model averaging method introduced by Hansen and Racine (2012) and Zhang et al. (2013), in which the model weights are determined by minimizing a leave-one-out cross-validation criterion. Hansen and Racine (2012) and Zhang et al. (2013) demonstrate that the JMA estimator is asymptotically optimal in the sense of achieving the lowest possible expected squared error.

(6) **Weighted-Average Least Squares (WALS)**: a new model averaging method, introduced by Magnus et al. (2010), which is a Bayesian combination of frequentist estimators and possesses both computational and theoretical advantages over frequentist and Bayesian methods (such as BMA, MMA and JMA). The computational advantage is that its computing time is linear in the number of predictor variables rather than exponential, as in frequentist and Bayesian model averaging techniques. The theoretical advantage is that, in contrast to standard BMA, which is based on normal priors leading to unbounded risk (prediction

variance), WALS is based on reflected Weibull, Subbotin, or Laplace priors, which can generate bounded risk.

The detail descriptions of all the above six sophisticated models are given in Appendix: *Sophisticated Forecast Models*.

The traditional way to estimate forecast models is based on recursively expanding windows or rolling windows. Following Pesaran and Timmermann (2007) and Pesaran and Pick (2011), we use the average windows forecasting method, which averages the same models but computed over different estimation windows instead of different models, to estimate forecast models. The AveW method is simple and attractive because no exact information about structural breaks is needed.

Given an observation window $\mathbb{W} = \{r_{t+1}, X_t\}_{t=0}^{T-1}$, we divide it into $m$ estimation windows

$$\mathbb{W}_i = \{r_{t+1}, X_t\}_{t=T-w_i}^{T-1}, \qquad i = 1, 2, ..., m$$

where $w_i = w_{min} + (\frac{i-1}{m-1})(T - w_{min})$ is the size of the $i$-th estimation window and $w_{min}$ is the size of a given minimum estimation window.

Then, the AveW forecast can be defined by the mean combination rule

不同时间区间预测值的平均值

$$\hat{r}_{T+1}^{AveW} = \frac{1}{m}\sum_{i=1}^{m} \hat{r}_{T+1}^{C}(\mathbb{W}_i) \tag{2}$$

where $\hat{r}_{T+1}^{C}(\mathbb{W}_i)$ is the forecast of WALS or any other models computed over the given estimation window $\mathbb{W}_i$. In recursive forecasting practice, we use the expanding windows as the input observation window $\mathbb{W}$ of the AveW method.

## *2.2 Combining Method*

In a similar spirit to Rapach et al. (2010)'s simple combining method for individual predictors and Lin et al. (2017)'s iterated combination approach that uses the historical average as shrinkage

target, we introduce a simple combination of the historical average with sophisticated models, defined as

$$\hat{r}^{C}_{T+1} = (1-\delta)\hat{r}^{HA}_{T+1} + \delta\hat{r}^{M}_{T+1} \tag{3}$$

where $\hat{r}^{HA}_{T+1}$ is the historical average forecast serving as the shrinkage target, $\hat{r}^{M}_{T+1}$ is the forecast of a sophisticated model (e.g., WALS) estimated using the average windows forecasting method, $\delta$ is the combination weight serving as the shrinkage factor.

Following Rapach et al. (2010), the combination weights can be set to equal weights or performance-based weights based on the forecasting performance (in terms of the MSFE) of individual models over a holdout out-of-sample period. The performance-based combination attaches greater weight to individual model with lower MSFE (better performance).[4]

There are two key differences between our simple combination method and the iterated combination approach. One is that, our method uses equal weights or performance-based weights, instead of the theoretically optimal weights that frequently do not perform well in practice (Timmermann, 2006; Rapach and Zhou, 2013). Second is that, to deal with parameter instability, our method estimates the sophisticated models using the average windows forecasting method, instead of the conventional expanding windows scheme like the iterated combination scheme.

By combining the simple but stable historical average forecast model with a sophisticated model (e.g. WALS) estimated using the average windows forecasting method, our approach is expected to improve forecasting performance in the same manner that asset diversification across a risk-free asset and a risky asset improves portfolio performance (see, Dunis et al., 2001; Timmermann, 2006; Rapach and Zhou, 2013, among others).

[4]Following Rapach et al. (2010), we always set the discount factor $\theta = 0.9$ when computing MSFE over the holdout out-of-sample period and set the holdout out-of-sample period as 120-month.

# 3 Data and Forecast Evaluation Measures

## *3.1 Data*

Following Rapach et al. (2010) and to make our results comparable with previous studies, we use updated monthly data from Welch and Goyal (2008) over the period 1926:12 to 2016:12.[5] Following Elliott et al. (2013) and to avoid multicollinearity when estimating some of the forecast combination models such as WALS, we consider only 12 of the 14 popular economic variables,[6] which can generate a total of $2^{12} = 4096$ candidate models. These 12 variables are as follows.

(1) Log dividend–price ratio [log(DP)]: log of a 12-month moving sum of dividends paid on the S&P 500 index minus the log of stock prices (S&P 500 index).

(2) Log dividend yield [log(DY)]: log of a 12-month moving sum of dividends minus the log of lagged stock prices.

(3) Log earnings–price ratio [log(EP)]: log of a 12-month moving sum of earnings on the S&P 500 index minus the log of stock prices.

(4) Stock variance (SVAR): monthly sum of squared daily returns on the S&P 500 index.

(5) Book-to-market ratio (BM): book-to-market value ratio for the Dow Jones Industrial Average.

(6) Net equity expansion (NTIS): ratio of a 12-month moving sum of net equity issues by New York Stock Exchange-listed stocks to the total end-of-year market capitalization of the New York Stock Exchange stocks.

(7) Treasury bill rate (TBL): interest rate on a three-month Treasury bill (secondary market).

[5]The data are available at http://www.hec.unil.ch/agoyal/. Variable definitions and data sources are described in more detail by Welch and Goyal (2008).

[6]Following Elliott et al. (2013), we exclude the log dividend–earnings ratio and the long-term yield, because the log dividend earnings ratio is equal to the difference between the log dividend price ratio and the log earnings–price ratio, while the long- term yield is equal to the sum of the term spread and the Treasury bill rate.

(8) Long-term return (LTR): return on long-term government bonds.

(9) Term spread (TMS): long-term yield minus the Treasury bill rate.

(10) Default yield spread (DFY): difference between BAA- and AAA-rated corporate bond yields.

(11) Default return spread (DFR): long-term corporate bond return minus the long-term government bond return.

(12) Inflation (INFL): calculated from the Consumer Price Index (all urban consumers).

The stock returns are measured as the difference between the log return on the S&P 500 (including dividends) and the log return on the risk-free Treasury bill. We follow Rapach and Zhou (2013) and divide the total sample into an in-sample period (1926:12–1956:12) and an out-of-sample evaluation period (1957:01–2016:12), including 720 observations. As the authors point out, the 1957:01–2016:12 forecast evaluation period covers most of the postwar era, including the oil price shocks of the 1970s; the deep recession associated with the Volcker disinflation in the early 1980s; the stock price shock in 1987; the long expansions of the 1960s, 1980s, and 1990s; and the recent global financial crisis in 2008 and the concomitant Great Recession. Following Rapach and Zhou (2013), we present not only the results of the full 1957:01–2016:12 forecast evaluation period, but also the results computed separately during National Bureau of Economic Research-dated business cycle expansions and recessions.

### *3.2 Forecast Evaluation Measures*

Following Rapach et al. (2010), we use the out-of-sample $R^2$ statistic, $R^2_{OS}$, suggested by Campbell and Thompson (2008) to evaluate the forecast accuracy of different (combination)

models. It is defined as

$$R^2_{OS} = 1 - \frac{MSFE^M}{MSFE^{bmk}} = 1 - \frac{\frac{1}{p}\sum_{t=T+1}^{T+p}(r_t - \hat{r}_t^M)^2}{\frac{1}{p}\sum_{t=T+1}^{T+p}(r_t - \bar{r}_t)^2} \tag{4}$$

where $[T+1, T+p]$ is the out-of-sample evaluation period, $MSFE^M$ is the MSFE of a forecast model $\mathcal{M}$ we investigate (such as the Kitchen sink, WALS, and BMA), $MSFE^{bmk}$ is the MSFE of the benchmark model.[7]

The out-of-sample $R^2_{OS}$ is a convenient statistic for comparing MSFEs and measures the proportional reduction in the MSFE for model $\mathcal{M}$ relative to the benchmark model. When $R^2_{OS} > 0$ , the forecast model $\mathcal{M}$ is more accurate than the benchmark in terms of the MSFE. The associated $p$-value is based on the work of Clark and West (2007) to test the null hypothesis that $R^2_{OS} \leq 0$.

Following Campbell and Thompson (2008), Welch and Goyal (2008) and Rapach et al. (2010), among others, we also analyze stock return forecasts with utility gains, which is a profit-based metric and provides more direct measures of the value of forecasts to a mean–variance investor who is more interested in the economic value of a forecast model than its precision.

Assume that a mean–variance investor with relative risk aversion parameter $\gamma$ will decide at the end of period $T$ to allocate the following share of a portfolio to equities in period $T+1$:

$$w_T^M = \frac{1}{\gamma}\left(\frac{\hat{r}_{T+1}^M}{\hat{\sigma}^2_{T+1}}\right) \tag{5}$$

where $\hat{r}^M_{T+1}$ is the forecast of model $\mathcal{M}$, and following Campbell and Thompson (2008) $\hat{\sigma}^2_{T+1}$ is the five-year rolling-window estimate of the variance of stock returns[8]. In addition, following Rapach et al. (2010) and Campbell and Thompson (2008), we constrain the portfolio weight on stocks to lie between 0% and 150% (inclusive), and set the relative risk aversion parameter $\gamma$ to three.

The investor then allocates $1 - w_T^M$ of the portfolio to risk-free bills, and the $T+1$ realized

[7] Following Rapach et al. (2010), we set the benchmark model to the historical average, defined as $\bar{r}_{T+1} = \frac{1}{T}\sum_{t=1}^{T} r_t$.
[8] The results are qualitatively similar for a ten-year moving window suggested by Rapach et al. (2016).

portfolio return is,

$$R_p^M = w_T^M \hat{r}_{T+1}^M + r_{T+1}^f \tag{6}$$

where $r_{T+1}^f$ is the risk-free rate. Then, the investor realizes a certainty equivalent return (CER) of the portfolio formed using the model $M$,

$$CER_M = \hat{u}_M - \frac{1}{2}\gamma(\hat{\sigma}_M^2) \tag{7}$$

where $\hat{u}_M$ and $\hat{\sigma}_M^2$ are the sample mean and variance, respectively, for the investor's portfolio over the evaluation period. Finally, in the case of monthly data, we obtain the utility (CER) gain in annualized percentage return as

$$\triangle(ann\%) = 1200 * (CER_M - CER_{bmk}) \tag{8}$$

where $CER_{bmk}$ is the CER of the portfolio formed using the benchmark model.

As pointed out by Rapach et al. (2010), the utility gain can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a forecast model relative to the information in the benchmark model alone.

In addition, we also calculate the monthly Sharpe ratio of the portfolio, which is the mean portfolio return in excess of the risk-free rate divided by the standard deviation of the excess portfolio return. To examine the adverse effect of transaction costs, we also consider the case of 50bps transaction costs, which is generally considered as a relatively high number.

## 4 Empirical Results

Before reporting the complete results for multivariate models, following Welch and Goyal (2008) and Rapach et al. (2010), we present the detailed forecasts based on the individual predictive regression models, in Table 1 for 1957:01–2016:12. It reinforces the findings of Welch and Goyal

(2008) that the individual predictive regression models cannot generate reliable out-of-sample forecasts of the equity premium. And, indeed, there is no single variable among the 12 considered that delivers a significantly positive $R^2_{OS}$ over the out-of-sample periods.

### *4.1 Performance Comparison of Sophisticated Models*

Table 2 provides the out-of-sample forecasting results of different forecast models, compared to Rapach et al. (2010)'s simple equal-weighted combination of individual predictors. Panel A indicates that, when estimated using recursively expanding windows, none of these sophisticated models including LASSO, Elastic net and WALS can significantly outperform the historical average in terms of the MSFE, although they can outperform the Kitchen sink model. It confirms that the simple equal-weighted combination forecast is superior.

One possible explanation for the relatively poor results under the conventional estimation scheme is that these sophisticated models (and Kitchen sink ) were developed for a stationary environment and cannot work well in environments with frequent structural changes, just as pointed out by Inoue et al. (2008). Now, we further investigate the performance of these models estimated using the AveW method, which has proved to be a useful tool for improving forecast accuracy in the presence of structural breaks. Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$.[9] We report results for the minimum estimation window size $w_{min} = 240$; results are qualitatively similar for other reasonable $w_{min}$ values.

Panel B of Table 2 provides the out-of-sample forecasting results of different models estimated using AveW. It shows that, compared to the conventional expanding windows scheme, AveW can modestly improve the performance (in terms of the MSFE) of all models we investigate. The $R^2_{OS}$ values of LASSO, Elastic net, JMA and WALS become positive at the 5% level of significance, meaning they outperform the historical average in terms of the MSFE. However, all these sophisticated models except LASSO and Elastic net still cannot outperform the simple

[9]Table A.1 in Internet Appendix shows that the out-of-sample performance of models estimated using AveW is quite robust to the number of estimation windows.

equal-weighted combination in terms of the MSFE. One possible explanation is that these sophisticated models without shrinkage suffer seriously from overfitting, as is the case with any highly parameterized model. The relative excellent performance of LASSO and Elastic net may result from their built-in shrinkage technique.

**[Place Table 2 about here]**

## *4.2 Results of Our Simple Combining Method*

Before reporting the complete results of our simple combining method for sophisticated models, which are estimated using the average window forecasting method, we present the results of the simple combinations of sophisticated models estimated traditionally, in Table A.2 in Internet Appendix. It shows that, when estimated using recursively expanding windows, equal-weighted combinations of the historical average with these sophisticated models still cannot achieve a significantly positive $R^2_{OS}$ over the out-of-sample periods, although they improve slightly (compared to the results in Panel A of Table 2).

We now investigate the benefits of our simple combining method, which allows for model uncertainty and parameter instability simultaneously, and as a result might work better.

### 4.2.1 The $R^2_{OS}$ Values

Table 3 reports the out-of-sample forecasting results of the simple equal-weighted combining method for individual predictors and different sophisticated models. It shows that, our method dramatically improves all the sophisticated models we investigate. All of these sophisticated models except BMA can outperform the simple equal-weighted combining method for individual predictors. For example, WALS achieves an impressive $R^2_{OS}$ value of 1.22% with a 5% level of significance, which is significantly larger than that of the combination of individual predictors. Surprisingly, all sophisticated models except BMA achieve a positive $R^2_{OS}$ over NBER-dated

business-cycle expansions, demonstrating a strong predictive ability in both good times and bad times. It is noteworthy that, under our scheme, even the standard Kitchen sink model demonstrates strong predictive ability.

**[Place Table 3 about here]**

Figure 1 depicts the relative performance of the equal-weighted combining method for individual predictors and different sophisticated models in terms of the DCSFE that is defined as the difference in the cumulative squared forecast error of the historical average benchmark model and the given model. Panel A illustrates that all the models except BMA consistently outperform the historical average over the entire evaluation period since 1970s. Under this scheme, also the performance of the Kitchen sink model becomes remarkably good, although it tends to be a bit volatile compared to WALS. At the same time, it seems that the performance of BMA becomes erratic after the stock price shock in 1987.

To investigate why there is such a substantial jump in the forecasting performance between 1987 and 1988, Panel B of Figure 1 further depicts the performance of different models during the period of 1987:01–1988:12.[10] It clearly shows that the performance of these models is quite stable before September 1987 and after November 1987. However, during the October 1987 stock market crash period, all the forecast models we investigate achieve a substantial leap, implying the predictive ability of macroeconomic variables to forecast extreme market movements particularly well. A point we address in more detail in Section 4.5.

**[Place Figure 1 about here]**

### 4.2.2 Asset Allocation Results

We further examine the economic value of the stock return predictability of the simple equal-weighted combining method for individual predictors and different sophisticated models from

[10] We thank the anonymous referee for suggesting discussing this issue.

an asset allocation perspective. Following Campbell and Thompson (2008) among others, we compute the certainty equivalent return (CER) gain and Sharpe ratio for a mean-variance investor who optimally allocates across equities and the risk-free asset using the out-of-sample predictive regression forecasts.

Table 4 reports the asset allocation results of different combinations of sophisticated models and individual predictors. It shows that, when assuming no transaction cost, WALS, MMA and even Kitchen sink achieve an annual utility gain of above 2%, about three times larger than that of the combination of individual predictors. The utility gains of Elastic net, JMA and LASSO are also quite high. Even if we assume 50bps transactions costs which is nowadays considered to be relatively high, all the sophisticated models we investigate can still achieve a positive utility gain.

**[Place Table 4 about here]**

In conclusion, our simple equal-weighted combining method can dramatically improve the performance (in terms of both the MSFE and utility gains) of the sophisticated models we investigate. When using performance-based combination weights, we can also draw a similar conclusion, see Table A.3 in Internet Appendix.

### *4.3 Forecast Encompassing Test Results*

To further assess the information content of the simple combinations of the historical average with different sophisticated models relative to the historical average or the simple combining method for individual predictors, following Rapach et al. (2010) among others, we conduct a forecast encompassing test. Harvey et al. (1998) develop the MHLN statistic for testing the null hypothesis that a given forecast contains all of the relevant information found in a competing forecast (i.e., the given forecast encompasses the competitor) against the alternative that the competing forecast contains relevant information beyond that in the given forecast.

Table 5 reports $p$-values of the Harvey et al. (1998)'s MHLN statistic for the historical average

forecast and different equal-weighted combinations, including the combination of individual predictors and the combinations of the historical average with different sophisticated models, such as WALS, MMA, BMA. Each entry in the table corresponds to a one-sided (upper-tail) test of the null hypothesis that the forecast of the model given in the column heading encompasses the forecast of the model given in the row heading against the alternative hypothesis that the forecast of the model given in the column heading does not encompass the forecast of the model given in the row heading. The $p$-values of the MHLN statistic are computed for the entire 1957:01–2016:12 out-of-sample forecast evaluation period.

The first column of Table 5 shows, not surprisingly, the historical average forecast fails to encompass the forecasts for any of the combinations of individual predictors and sophisticated models (except BMA). The second column shows that the combination of individual predictors significantly encompasses the historical average, but fails to encompass the forecasts for WALS and Kitchen sink. Interestingly, the third column indicates that Kitchen sink contains incremental forecasting information beyond all other models we investigate, including WALS and the combination of individual predictors. The fourth column shows that, compared to Kitchen sink, WALS always contains more incremental forecasting information beyond all other models we investigate, suggesting potential gains from WALS. From the fifth to eighth column, we can draw a similar conclusion that MMA, JMA, Elastic net and LASSO contain incremental forecasting information beyond all other models we investigate, with the exception that MMA fails to encompass WALS. The ninth column of Table 5 shows that BMA cannot encompass any other sophisticated models, reinforcing the conclusion that the performance of BMA is erratic. However, it still contains incremental forecasting information beyond the historical average and the combination of individual predictors.

**[Place Table 5 about here]**

### *4.4 Why Shrinkage Can Help?*

We now investigate the benefits of shrinkage in improving forecasting performance, based on the well-known bias-variance decomposition expressed as

$$MSFE(\hat{r}^c_{t+1}) = Bias(\hat{r}^c_{t+1})^2 + Variance(\hat{r}^c_{t+1}) \tag{9}$$

where $Variance(\hat{r}^c_{t+1}) = E[E(\hat{r}^c_{t+1}) - \hat{r}^c_{t+1}]^2$ and $Bias(\hat{r}^c_{t+1}) = E(\hat{r}^c_{t+1}) - r_{t+1}$.

By explicitly adjusting the shrinkage factor $\delta$ in Equation (3), our combining scheme provides the means for analyzing the effects of shrinkage. We consider a grid of values of $\delta \in \{0.01, 0.02, ..., 0.99, 1.00\}$ to allow for different degrees of shrinkage.

Figure 2 plots the bias-variance decomposition of the $MSFE(\hat{r}^c_{t+1})$ of the equal-weighted combination of the historical average with Kitchen sink, against the shrinkage factor $\delta$.[11] It demonstrates that, the $MSFE(\hat{r}^c_{t+1})$, $Variance(\hat{r}^c_{t+1})$ and $Bias(\hat{r}^c_{t+1})^2$ decrease dramatically with the increase of the shrinkage factor $\delta$, especially when $\delta \leq 0.38$ in this experiments. After that, the $MSFE(\hat{r}^c_{t+1})$ and $Variance(\hat{r}^c_{t+1})$ gradually increase while $Bias(\hat{r}^c_{t+1})^2$ continues to decrease. It clearly shows that the $MSFE(\hat{r}^c_{t+1})$ is mainly determined by the $Variance(\hat{r}^c_{t+1})$ because $Bias(\hat{r}^c_{t+1})^2$ is so small that its effect on the $MSFE(\hat{r}^c_{t+1})$ is negligible.

It is natural to ask why the variance of the combination forecast can be smaller than that of individual models such as HA and Kitchen sink. Without losing generality, looking back to the simple equal-weighted combination of HA with Kitchen sink $\hat{r}^c_{t+1} = \frac{1}{2}(\hat{r}^{HA}_{t+1} + \hat{r}^{KS}_{t+1})$, its forecast variance can be further decomposed into three components

$$\begin{aligned} Variance(\hat{r}^c_{t+1}) &= E[E(\hat{r}^c_{t+1}) - \hat{r}^c_{t+1}]^2 = E[\frac{1}{2}(E(\hat{r}^{HA}_{t+1}) - \hat{r}^{HA}_{t+1}) + (E(\hat{r}^{KS}_{t+1}) - \hat{r}^{KS}_{t+1})]^2 \\ &= \frac{1}{4}[Variance(\hat{r}^{HA}_{t+1}) + Variance(\hat{r}^{KS}_{t+1}) + 2Covariance(\hat{r}^{HA}_{t+1}, \hat{r}^{KS}_{t+1})] \end{aligned} \tag{10}$$

[11] For brevity, we only report the results of Kitchen sink. Under our simple combining scheme, the conclusions for other sophisticated models we investigate are similar.

where $Covariance(\hat{r}^{HA}_{t+1}, \hat{r}^{KS}_{t+1}) = E[(\hat{r}^{HA}_{t+1}) - \hat{r}^{HA}_{t+1})(\hat{r}^{KS}_{t+1}) - \hat{r}^{KS}_{t+1})]$.

It indicates that $Variance(\hat{r}^{c}_{t+1})$ depends heavily on the covariance term that can be negative, while $Variance(\hat{r}^{HA}_{t+1})$ and $Variance(\hat{r}^{KS}_{t+1})$ are constrained to be positive. It is obvious that we can reduce the variance of the combination forecast by carefully choosing a shrinkage target (e.g. the historical average) that is not perfectly positively correlated with the sophisticated models (e.g. Kitchen sink or WALS). The reduction of variance results into a smaller MSFE.

In fact, like other forecast combination methods, our simple combination scheme also improves forecasting performance in the same manner that asset diversification across a risk-free asset and a risky asset improves portfolio performance, just as pointed out by Dunis et al. (2001); Timmermann (2006); Rapach and Zhou (2013); Lin et al. (2017) among others.

Furthermore, Figure 2 confirms that the optimal shrinkage (combination weight) is hard to estimate, just as pointed out by (Timmermann, 2006; Rapach and Zhou, 2013). It shows that, while the true optimal shrinkage is 0.38, the in-sample estimation of the optimal shrinkage based on the method proposed by Lin et al. (2017) is 0.21. However, it also demonstrates that, by always specifying shrinkage $\delta = 0.5$, our simple combining scheme achieves an acceptable MSFE.

**[Place Figure 2 about here]**

## *4.5 Forecasting Extreme Market Movements*

Fama and French (1989) and Campbell and Cochrane (1999); Cochrane (2007) argue that heightened risk aversion during economic downturns demands a higher risk premium, thereby generating equity premium predictability. Following Rapach et al. (2010), we examine the $R^2_{OS}$ statistics computed separately during NBER-dated business-cycle recessions (bad times) and expansions (good times). Surprisingly, our results in Table 3 show that the predictive ability of almost all the simple equal-weighted combinations of sophisticated models we investigate remains strong in both bad times and good times.

In order to explore the economic sources of equity premium predictability, we next examine forecasting gains of the simple equal-weighted combining method for sophisticated models and individual predictors during extreme (downturn) periods of normalized monthly returns $r_t$. Panel A of Table 6 reports the results of the combinations of WALS, Kitchen sink and individual predictors for four extreme periods, defined as $|r_t| \geq 0.5, 1.0, 1.5, 2.0$ respectively. It shows that the $R^2_{OS}$ statistics of these combinations are always higher during extreme periods than during less extreme periods. For example, the $R^2_{OS}$ statistics are almost two times higher during the extreme period when $|r_t| \geq 2.0$ than during the extreme period when $|r_t| \geq 0.5$. Surprisingly, under our combination scheme, the $R^2_{OS}$ statistics of Kitchen sink always beat WALS during extreme periods, although it underperforms WALS over the full evaluation period.

Panel B reports the results for four extreme downturn periods, defined as $r_t \leq -0.5$, -1.0, -1.5, -2.0 respectively. It shows that these combinations always achieve much higher $R^2_{OS}$ statistics during extreme downturn periods than during the corresponding extreme periods that also include extreme upturn periods. For example, the $R^2_{OS}$ statistics of the combination of Kitchen sink are more than three times higher during the extreme downturn period when $r_t \leq -0.5$ than during the extreme period when $|r_t| \geq 0.5$.[12]

Overall, Table 6 demonstrates that enhanced forecasting gains for the simple equal-weighted combinations of sophisticated models mainly stem from extreme periods, especially extreme downturn periods. It also shows that the combinations of the historical average with WALS or Kitchen sink always achieve much higher $R^2_{OS}$ statistics than the combination of individual predictors during extreme (downturn) periods. It confirms that the behavior of these simple combination forecasts agrees with the Fama and French (1989) and Campbell and Cochrane (1999); Cochrane (2007) account of equity premium predictability.

**[Place Table 6 about here]**

[12] In fact, our untabulated results further indicate that out-of-sample gains for forecasts of all models we investigate are mainly concentrated in downturn periods defined as $r_t < 0$, especially extreme downturn periods.

## 5 Forecasting Characteristics Portfolios

To verify the effectiveness of our simple combining method for sophisticated models, we next explore the return predictability of characteristic portfolios, instead of the market portfolio S&P500, with the same set of 12 economic variables. These tests not only strengthen our previous findings for aggregate stock market predictability, but also enhance our understanding of the economic channels through which economic variables impact asset prices in the presence of model uncertainty and parameter instability.

We consider the well-known 10 size characteristics portfolios available from Kenneth French's data library. The monthly returns of these portfolios are value-weighted. The out-of-sample evaluation period is 1957:01–2016:12. For brevity, we only report the results of Kitchen sink. We expect that Kitchen sink with the same set of economic variables can also present strong forecasting power for returns of these characteristic portfolios. For comparison, the results of the simple equal-weighted combining method for individual predictors are also reported.

Table 7 reports the $R^2_{OS}$ values of these simple equal-weighted combinations for different characteristic portfolios. It shows that both of the two combinations have positive $R^2_{OS}$ values for all the 10 size characteristic portfolios. Not surprisingly, the combination of Kitchen sink can significantly outperform the combination of individual predictors in terms of the MSFE, utility gain and Sharpe ratio.

**[Place Table 7 about here]**

## 6 Forecasting Macro Conditions

In order to explore the economic sources of equity premium predictability, we investigate whether the stock return predictors also have predictive power for future business conditions. As pointed out by Cochrane (2008), if the return predictor shows predictive power for business cycle,

then the predictable return variations are more plausibly related to macroeconomic risk.

In particular, we run the following regressions with the simple equal-weighted combining method,

$$Y_{t+1} = \alpha + X_t\beta + e_{t+1} \tag{11}$$

$$Y_{t+12} = \alpha + X_t\beta + e_{t+1} \tag{12}$$

$$\Delta Y_{t+1} = \alpha + X_t\beta + e_{t+1} \tag{13}$$

$$\Delta Y_{t+12} = \alpha + X_t\beta + e_{t+1} \tag{14}$$

where $X_t$ is the same 12 economic predictors, $Y_{t+1}$ is the macroeconomic condition for next month, $Y_{t+12}$ is the macroeconomic condition for next year, $\Delta Y_{t+1} = Y_{t+1} - Y_{(t+1)-12}$ is the year change in macroeconomic condition for the next month, and $\Delta Y_{t+12} = Y_{t+12} - Y_t$ is the year change in macroeconomic condition for the next year.

We focus on the following macroeconomic conditions $Y_t$

(1) Chicago Fed National Activity Index (CFNAI). The CFNAI is a monthly index designed to capture economic activty and inflationary presure. The data are download from the Federal Reserve Bank of Chicago. Data spans from 1967:03 to 2017:12, monthly.

(2) Smoothed U.S. Recession Probabilities (SRP). Smoothed recession probabilities for the United States are obtained from a dynamic-factor markov-switching model applied to four monthly coincident variables: non-farm payroll employment, the index of industrial production, real personal income excluding transfer payments, and real manufacturing and trade sales. The data are download from the Federal Reserve Bank of St. Louis. Data spans from 1967:06 to 2017:12, monthly.

(3) Industrial Production Growth (IPG). The production growth rate data are also obtained from the Federal Reserve Bank of St. Louis. Data spans from 1919:01 to 2017:12, monthly.

(4) Macroeconomic Uncertainty Index (MU). MU is the macroeconomic uncertainty index proposed in Jurado et al. (2015), which is constructed as a common component of the unpredictable variation of macroeconomic variables. It can be download from Prof. Ludvigson's website. Data spans from 1960:07 to 2017:12, monthly.

(5) Output Gap (Gap): The output gap is the difference between actual GDP or actual output and potential GDP. Both of them can be obtained from Federal Reserve Bank of St. Louis. Data spans from 1968:01 to 2017:12, monthly.

(6) Cay: Cay, introduced by Lettau and Ludvigson (2001a,b), is a cointegrating residual between log consumption, log asset (nonhuman) wealth, and log labor income. It is proved to have striking forecasting power for excess returns on aggregate stock market indexes. It can be download from Prof. Ludvigson's website. Data spans from 1952:Q1 to 2017:Q3, quartly. We tranform it to monthly data spanning from 1952:01 to 2017:12.

(7) Civilian Unemployment Rate (UNRATE). The UNRATE is also obtained from Federal Reserve Bank of St. Louis. Data spans from 1948:01 to 2017:12, monthly.

Table 8 reports the $R^2_{OS}$ values of forecasts of the simple equal-weighted combinations of the historical average with Kitchen sink for the above macroeconomic condition variables. It shows that, not surprisingly, all the out-of-sample $R^2_{OS}$ values of these combinations are significantly positive, indicating that the 12 economic predictors also have strong predictive power for future business conditions.

**[Place Table 8 about here]**

## 7 Forecasting with an Alternative Dataset

As a robustness check of our simple combining method for sophisticated models, we next explore the return predictability of S&P 500 index based on factors extracted from FRED-MD, a

macroeconomic database of 134 monthly U.S. indicators.[13] The data spans from 1959:01 through 2016:12. Mccracken and Ng (2015) shows that factors extracted from the FRED-MD dataset can potentially be useful for predicting macroeconomic aggregates, such as U.S. industrial production, nonfarm employment, headline CPI inflation, and core CPI inflation.

Following Mccracken and Ng (2015) and Stock and Watson (2002), we estimate the static factors by principal component analysis adapted to allow for missing values. The number of factors $K = 7$ is determined by an information criterion based method introduced by Bai and Ng (2002).[14]

We present the out-of-sample forecasting results of different forecast models with factors extracted from FRED-MD, in Table A.4 in Internet appendix for 1980:01–2016:12. Panel A indicates that, when estimated using recursively expanding windows, none of these sophisticated models can outperform the simple equal-weighted combination in terms of the MSFE, although LASSO, Elastic net and BMA can slightly outperform the historical average. It also reinforces that the FRED-MD dataset contains predictive content. Panel B shows that, when estimated using the average window forecasting method, all models we investigate (except Kitchen sink) can achieve a significant positive $R^2_{OS}$. LASSO and BMA can even outperform the simple equal-weighted combination.

Table 9 reports the $R^2_{OS}$ values of the out-of-sample forecasts of the equal-weighted combining method for individual predictors and different sophisticated models based on factors extracted from the FRED-MD dataset. It shows that, based on our combining method, all forecast models we investigate can be further improved substantially. All forecast models including the Kitchen sink model can outperform the simple equal-weighted combination of individual predictors.

**[Place Table 9 about here]**

[13]The dataset is available at https://research.stlouisfed.org/econ/mccracken/sel/. Variable definitions and data sources are described in more detail by Mccracken and Ng (2015).

[14]We thank McCracken for providing the Matlab code to estimate and select factors.

# 8 Conclusions

By introducing a simple combining method, we compare several representative sophisticated models that deal with model uncertainty, such as the recently proposed WALS, the popular BMA, the famous LASSO and Elastic net, the recent MMA and JMA, the standard Kitchen sink model, in the context of stock return forecasting based on macroeconomic variables.

Our results confirm, when estimated conventionally, these sophisticated models we investigate perform poorly and the simple combination of individual predictors is superior. However, if they are used in a shrinkage fashion and estimated using the average windows forecasting method as in our simple combination scheme, all these models improve substantially and can even outperform the simple combination of individual predictors in terms of both the MSFE and utility gains. For example, the equal-weighted combination of the historical average with Kitchen sink achieves a statistically significant monthly out-of-sample $R^2_{OS}$ of 1.10% and utility gain of 2.34%, both of which are much higher than that of the simple combination of individual predictors.

Further investigation shows that out-of-sample forecasting gains of our scheme are mainly concentrated in extreme market downturns, especially during extreme market downturns. It confirms that the behavior of these combination forecasts agrees with the common view of the literature that heightened risk aversion during economic downturns requires a higher risk premium.

The good performance of our combination scheme can be explained by the Modern Portfolio Theory. The key to its success stems from diversification across the historical average model with near-zero variance and a sophisticated model with much higher variance, just like diversification across a risk-free asset and risky assets that improves portfolio performance.

While our empirical results suggest that our approach with equal-weighted or performance-based combination weights can considerably improve stock return forecasting, how to determine the optimal combination weights remains an open question. The extensive research results of portfolio optimization in the literature may provide insights into the further study on this question.

# Appendix: Sophisticated Forecast Models

## *Variable Selection Method: LASSO and Elastic Net*

The LASSO algorithm introduced by Tibshirani (1996) is a popular regularization method for performing shrinkage in regressions. It can improve the prediction accuracy and interpretability of regression models by altering the model fitting process to select only a subset of the provided variables for use in the final model rather than using all of them.

For the standard regression model, Equation (1), the objective of LASSO can be expressed as

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^N}{\operatorname{argmin}} \left( \frac{1}{T} \sum_{t=0}^{T-1} (r_{t+1} - X_t \beta)^2 + \lambda \sum_{i=1}^{N} |\beta_i| \right) \tag{A.1}$$

where $\lambda$ is a regularization parameter. The first component in parentheses is the familiar sum of squared residuals, so that the objective function reduces to that for the standard regression model when $\lambda = 0$. The second component is an $\iota_1$ penalty term that shrinks the slope coefficient estimates to prevent overfitting. Based on the lasso method in Matlab, We use 5-fold cross-validation to choose the optimal $\lambda$.

To overcome some limitations of LASSO, the Elastic net algorithm introduced by Zou and Hastie (2005) adds an $\iota_2$ penalty, expressed as

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^N}{\operatorname{argmin}} \left( \frac{1}{T} \sum_{t=0}^{T-1} (r_{t+1} - X_t \beta)^2 + \lambda \sum_{i=1}^{N} \left( \frac{1-\alpha}{2} \beta_i^2 + \alpha |\beta_i| \right) \right) \tag{A.2}$$

When $\alpha = 1$, Elastic net is the same as LASSO. Given a fixed $\alpha = 0.5$, we use the LASSO method with 5-fold cross-validation in Matlab to obtain the optimal $\lambda$ for the Elastic net algorithm.

## *Model Averaging Techniques*

Let $\lambda = (\lambda_1, ..., \lambda_m)$ be a vector of model weights with $0 \leq \lambda_i \leq 1$ and $\sum_{i=1}^{m} \lambda_i = 1$ where $m = 2^N$ is the number of candidate models and $\lambda_i$ is the weight of model $\mathcal{M}_i$. A general forecast combination can be expressed as

$$\hat{r}_{T+1}^{C} = \sum_{i=1}^{m} \lambda_i \hat{r}_{T+1}^{(i)} = X_T \hat{\beta}(\lambda) \tag{A.3}$$

where $\hat{r}_{T+1}^{(i)}$ is the forecast result of model $\mathcal{M}_i$ and $\hat{\beta}(\lambda)$ is an averaging estimator of $\beta$.

### Bayesian Model Averaging

A natural framework for forecast combination is offered by the popular Bayesian model averaging, where the weights of all candidate models are determined by their posterior probabilities (see Raftery et al., 1997; Hoeting et al., 1999; Avramov, 2002).

Let $P(\mathcal{M}_i)$ denote the prior probability that $\mathcal{M}_i$ is a true model and $P(R|\mathcal{M}_i)$ denote the marginal likelihood of $R$ in model $\mathcal{M}_i$ . Then, the BMA weight for $\mathcal{M}_i$ is given by

$$\lambda_i^{BMA} = P(\mathcal{M}_i|R) = \frac{P(\mathcal{M}_i)P(R|\mathcal{M}_i)}{\sum_{j=1}^{m} P(\mathcal{M}_j)P(R|\mathcal{M}_j)}, \quad i = 1, 2, ..., m \tag{A.4}$$

Under the assumption of equal model priors and diffuse model priors on parameters, Buckland et al. (1997) provide a good approximation for the BMA weights, Equation (A.4), expressed as

$$\lambda_i^{BMA} = exp(-\frac{1}{2}BIC_i) / \sum_{j=1}^{m} exp(-\frac{1}{2}BIC_j), \quad i = 1, 2, ..., m \tag{A.5}$$

where $BIC_i = Tlog(\hat{\sigma}_i^2) + log(T)(N_i)$ is the Bayesian information criterion (BIC) for model $\mathcal{M}_i$, $\hat{\sigma}_i^2$ is the estimate of $\sigma^2$ in $\mathcal{M}_i$ and $N_i$ is the actual number of predictors of $\mathcal{M}_i$.

Then, based on Equations (A.3) and (A.5), the BMA forecast model can be expressed as

$$\hat{r}_{T+1}^{BMA} = \sum_{i=1}^{m} \lambda_i^{BMA} \hat{r}_{T+1}^{(i)} \tag{A.6}$$

BMA is straightforward but has several problems in practice. The most important is that the computation time will increase exponentially with the increase of $N$. The second is that, since the priors are based on the normal distribution, they can lead to unbounded prediction variance.

**Frequentist Model Averaging**

In contrast to BMA which rely on priors, frequentist model averaging (FMA) techniques require no priors and the corresponding model weights are fully determined by the data. Therefore, FMA approaches have received much attention and have made significant progress over the last decades. Based on different weighting schemes, many kinds of FMA approaches exist.

**Mallows Model Averaging** Hansen (2007) and Wan et al. (2010) have developed the Mallows model averaging method for homoskedastic linear regression models, in which the model weights are selected by minimizing a Mallows criterion. Hansen (2008) shows that the Mallows criterion is an asymptotically unbiased estimator of both the in-sample mean squared error and the out-of-sample one-step-ahead MSFE.

Let $K = (K_1, ..., K_m)$ where $K_i$ denotes the number of variables in model $\mathcal{M}_i$. Define the Mallows criterion as

$$C_p(\lambda) = (R - X\hat{\beta}(\lambda))'(R - X\hat{\beta}(\lambda)) + 2\sigma^2 \lambda K \tag{A.7}$$

where $\lambda$ and $\hat{\beta}(\lambda)$ are defined in Equation (A.3).

Then, by minimizing this criterion, we find the selected weight vector to be

$$\lambda^{MMA} = \arg\min_{\lambda \in \mathcal{H}} C_p(\lambda) \tag{A.8}$$

where $\mathcal{H} = \{\lambda \in \mathbb{R}^m : \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1\}$.

Based on Equations (A.3) and (A.8), the MMA forecast model can be expressed as

$$\hat{r}_{T+1}^{MMA} = \sum_{i=1}^{m} \lambda_i^{MMA} \hat{r}_{T+1}^{(i)} \tag{A.9}$$

**Jackknife Model Averaging** Hansen and Racine (2012) and Zhang et al. (2013) develop the jackknife model averaging method to select the weights of heteroskedastic linear regression models by minimizing a leave-one-out cross-validation criterion. They demonstrate that the JMA estimator is asymptotically optimal in the sense of achieving the lowest possible expected squared error.

Let $\tilde{R}^{(i)} = (\tilde{r}_1^{(i)}, ..., \tilde{r}_T^{(i)})$ be the jackknife estimator of $R$ in the following model $\mathcal{M}_i$:

$$\tilde{r}_t^{(i)} = X_t \hat{\beta}_{(-t)}$$

where $\hat{\beta}_{(-t)}$ is the OLS estimator of $\beta$ computed with the $t$-th observations deleted.

The jackknife version of the averaging estimator is

$$\tilde{R}(\lambda) = \sum_{i=1}^{m} \lambda_i \tilde{R}^{(i)}$$

The jackknife version of the averaging residual is

$$\tilde{e}(\lambda) = R - \tilde{R}(\lambda) = \sum_{i=1}^{m} \lambda_i \tilde{e}^{(i)} = \lambda \tilde{e}$$

where $\tilde{e} := (\tilde{e}^{(1)}, ..., \tilde{e}^{(m)})$ , and $\tilde{e}^{(i)} = R - \tilde{R}^{(i)}$.

Define the leave-one-out cross-validation criterion as

$$CV_T(\lambda) = \frac{1}{T} \tilde{e}(\lambda)' \tilde{e}(\lambda) \tag{A.10}$$

Then, by minimizing this criterion, we find the selected weight vector to be

$$\lambda^{JMA} = \arg\min_{\lambda \in \mathcal{H}} CV_T(\lambda) \tag{A.11}$$

where $\mathcal{H} = \{\lambda \in \mathbb{R}^m : \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1\}$.

Based on Equations (A.3) and (A.11), the JMA forecast model can be expressed as

$$\hat{r}_{T+1}^{JMA} = \sum_{i=1}^{m} \lambda_i^{JMA} \hat{r}_{T+1}^{(i)} \tag{A.12}$$

MMA and JMA is computationally very demanding when the model space is large. Following Hansen (2007), when estimating MMA and JMA, we use the restriction of pure nested models, in which variables are ordered using the LASSO method in Matlab.

**Weighted-Average Least Squares**

Weighted-average least squares introduced by Magnus et al. (2010) is a Bayesian combination of frequentist estimators, which possesses both computational and theoretical advantages over frequentist and Bayesian methods. The computational advantage is that its computing time is linear in the number of predictor variables rather than exponential, as in Bayesian model averaging techniques. The theoretical advantage is that, in contrast to standard BMA, which is based on normal priors leading to unbounded prediction variance, WALS is based on reflected Weibull, Subbotin, or Laplace priors, which imply a coherent treatment of ignorance and can generate bounded prediction variance.

Rewrite the standard predictive regression model, Equation (1), in matrix form as

$$R = X_1\beta_1 + X_2\beta_2 + e \tag{A.13}$$

where $X_1$ is a $T \times (N_1 + 1)$ vector of "focus" predictors that must be included in the model based

on theoretical or other grounds and always includes a constant term as the first variable, $X_2$ is a $T \times N_2$ vector of "auxiliary" predictors that may or may not be included in the model.

Let $P$ be an orthogonal matrix and $\Lambda$ a diagonal matrix with positive diagonal elements such that $P'X_2'M_1X_2P = \Lambda$, where

$$M_1 = I_T - X_1(X_1'X_1)^{-1}X_1' \tag{A.14}$$

and $I_T$ represents a $T \times T$ identity matrix.

Define the transformed auxiliary variables and parameter as

$$X_2^* = X_2P\Lambda^{-1/2}, \qquad \beta_2^* = \Lambda^{1/2}P'\beta_2 \tag{A.15}$$

such that $X_2^*\beta_2^* = X_2\beta_2$.

Then we can rewrite Equation (A.13) equivalently as

$$R = X_1\beta_1 + X_2^*\beta_2^* + \mathbf{e} \tag{A.16}$$

This transformation is called semi-orthogonal because the new matrix $(X_1 : X_2^*)$ is semi-orthogonal in the sense that $X_2^*M_1X_2^{*'} = I_{N_2}$.

The equivalence theorem (Magnus et al., 2010, 2016) tells us that the WALS estimators of $\beta_1$ and $\beta_2^* = (\beta_{2,1}^*, \beta_{2,2}^*, ..., \beta_{2,N_2}^*)$ in Equation (A.16) can be expressed as, respectively,

$$\hat{\beta}_1 = \hat{\beta}_{1r} - Q^*\hat{\beta}_2^* \tag{A.17}$$

$$\hat{\beta}_{2,h}^* = \hat{\beta}_{2u,h}^* - \hat{\sigma}_h \frac{A_1(\frac{\hat{\beta}_{2u,h}^*}{\hat{\sigma}_h})}{A_0(\frac{\hat{\beta}_{2u,h}^*}{\hat{\sigma}_h})}, \quad (h = 1, ..., N_2) \tag{A.18}$$

where

$$\hat{\beta}_{1r} = (X_1'X_1)^{-1}X_1'R$$
$$Q^* = (X_1'X_1)^{-1}X_1'X_2^*$$
$$\hat{\beta}_{2u}^* = X_2^{*'}M_1R$$
$$A_j(x) = \int_{-\infty}^{\infty}(x-\gamma)^j\phi(x-\gamma)\pi(\gamma)d\gamma, \quad (j=0,1)$$

, $\phi$ denotes the standard normal density, and $\pi(\gamma)$ is the reflected Weibbull prior[15], and $\hat{\sigma}_h$ is the estimate of the standard deviation of $\beta_{2,h}^*$.

Based on Equations (A.16), (A.17) and (A.18), we can easily obtain the WALS forecast, as follows:

$$\hat{r}_{T+1}^{WALS} = X_{1,T}\hat{\beta}_1 + X_{2,T}^*\hat{\beta}_2^* \tag{A.19}$$

In practice, we consider the implementation of WALS with the constant term as the only one focus variable, meaning that we are not sure which variable should be included in the model.

The WALS theory is appealing and has turned out to be an effective approach for dealing with model uncertainty. Fox example, Magnus et al. (2016) compare it with four competing predictors (unrestricted model, pretesting, ridge regression, MMA) in a wide range of simulation experiments. They find that the WALS predictor generally produces the lowest mean squared error. They also find that the estimated variance of the WALS predictor is typically larger than the variance of the pretesting and ridge predictor but more accurate in terms of the root mean squared error. Finally, when model uncertainty increases, the dominance of WALS becomes more pronounced.

[15]Following Magnus and Luca (2016), the reflected Weibull prior is preferred, which is defined as

$$\pi(\gamma) = \frac{qc}{2}|\gamma|^{-(1-q)}exp^{-c|\gamma|^q}$$

, where $q = 0.8876$ , $c = log2$.

## References


Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64 (3), 423–458.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

Buckland, S. T., Burnham, K. P., Augustin, N. H., 1997. Model selection: An integral part of inference. Biometrics 53 (2), 603–618.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumptionbased explanation of aggregate stock market behavior. Journal of Political Economy 107 (2), 205–251.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Scholarly Articles 21 (4), 1509–1531.

Claeskens, G., Jansen, M., 2008. Model selection and model averaging. International Encyclopedia of the Social and Behavioral Sciences 172 (4), 647–652.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138 (1), 291–311.

Cochrane, J. H., 2007. Financial markets and the real economy. in R. Mehra (ed.), Handbook of the Equity Premium. Amsterdam: Elsevier.

Cochrane, J. H., 2008. Chapter 7 financial markets and the real economy. Handbook of the Equity Risk Premium (January), 237–325.

Dunis, C., Moody, J., Timmermann, A., 2001. Developments in forecast combination and portfolio choice. International Journal of Forecasting 18 (3), 462–463.

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. Journal of Econometrics 177 (2), 357–373.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds . Journal of Financial Economics 25 (1), 23–49.

Hansen, B. E., 2007. Least squares model averaging. Econometrica 75 (4), 1175–1189.

Hansen, B. E., 2008. Least-squares forecast averaging. Journal of Econometrics 146 (2), 342–350.

Hansen, B. E., Racine, J. S., 2012. Jackknife model averaging. Journal of Econometrics 167 (1), 38–46.

Harvey, D. I., Leybourne, S. J., Newbold, P., 1998. Tests for forecast encompassing. Journal of Business & Economic Statistics 16 (2), 254–259.

Hoeting, J. A., Madigan, D., Raftery, A. E., Volinsky, C. T., 1999. Bayesian model averaging: A tutorial. Statistical Science 14 (4), 382–401.

Inoue, Atsushi, Kilian, Lutz, 2008. How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. Journal of the American Statistical Association 103 (482), 511–522.

Jurado, K., Ludvigson, S. C., Ng, S., March 2015. Measuring uncertainty. American Economic Review 105 (3), 1177–1216.

Lettau, M., Ludvigson, S., 2001a. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56 (3), 815–849.

Lettau, M., Ludvigson, S., 2001b. Resurrecting the (c)capm: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109 (6), 1238–1287.

Lin, H., Wu, C., Zhou, G., 2017. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science.

Magnus, J. R., Luca, G. D., 2016. Weighted-average least squares (WALS): A survey. Journal of Economic Surveys 30 (1), 117–148.

Magnus, J. R., Powell, O., Prfer, P., 2010. A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics 154 (2), 139–153.

Magnus, J. R., Wang, W., Zhang, X., 2016. Weighted-average least squares prediction. Econometric Reviews 35 (6), 1040–1074.

Mccracken, M. W., Ng, S., 2015. Fred-md: A monthly database for macroeconomic research. Journal of Business and Economic Statistics 34 (4).

Pesaran, M. H., Pick, A., 2011. Forecast combination across estimation windows. Journal of Business and Economic Statistics 29 (2), 307–318.

Pesaran, M. H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. Journal of Finance 50 (4), 1201–1228.

Pesaran, M. H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137 (1), 134–161.

Raftery, A. E., Madigan, D., Hoeting, J. A., 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92 (437), 179–191.

Rapach, D., Zhou, G., 2013. Forecasting stock returns. In: Handbook of Economic Forecasting. Elsevier B.V., Ch. 6, pp. 328–383.

Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns . Journal of Financial Economics 121 (1), 46–65.

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

Stock, J. H., Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14 (1), 11–30.

Stock, J. H., Watson, M. W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20 (2), 147–162.

Stock, J. H., Watson, M. W., 2006. Forecasting with many predictors. In: Handbook of Economic Forecasting. Vol. 1. Elsevier B.V., Ch. 10, pp. 515–554.

Tibshirani, R. J., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society.Series B (methodological). Wiley 58, 267–288.

Timmermann, A., 2006. Chapter 4 forecast combinations. Vol. 1 of Handbook of Economic Forecasting. Elsevier, pp. 135 – 196.

Wan, A. T. K., Zhang, X., Zou, G., 2010. Least squares model averaging by mallows criterion. Journal of Econometrics 156 (2), 277–283.

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

Zhang, X., Wan, A. T. K., Zou, G., 2013. Model averaging by jackknife criterion in models with dependent data. Journal of Econometrics 174 (2), 82–94.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 67 (2), 301–320.

**Table 1**
**Monthly U.S. equity premium out-of-sample forecasting results of individual economic variables, 1957:01–2016:12**

This table reports the $R^2_{OS}$ values of forecast models based on individual economic variables estimated using recursively expanding windows. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the forecast model given in the first column relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. $R^2_{OS,exp}$ ($R^2_{OS,rec}$) statistics are calculated over NBER-dated business-cycle expansions (recessions). The out-of-sample evaluation period is 1957:01–2016:12.

| Economic variable | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ |
|---|---|---|---|---|
| log(DP) | -0.05 | 1.31 | -1.24 | 2.41 |
| log(DY) | -0.37 | 1.49* | -2.28 | 3.56 |
| log(EP) | -1.88 | 0.57 | -2.21 | -1.20 |
| SVAR | 0.32 | 0.97 | -0.02 | 1.01 |
| BM | -1.74 | 0.50 | -2.56 | -0.04 |
| NTIS | -0.91 | 0.22 | 0.50 | -3.82 |
| TBL | -0.01 | 1.34* | -0.84 | 1.71 |
| LTR | -0.08 | 1.01 | -0.40 | 1.00 |
| TMS | 0.06 | 0.84 | -0.85 | 1.52 |
| DFY | -0.04 | -0.24 | -0.06 | -0.01 |
| DFR | -0.01 | 0.31 | 0.12 | -0.28 |
| INFL | -0.09 | 0.01 | 0.10 | -0.48 |

**Table 2**
**Out-of-sample performance of sophisticated models estimated traditionally**

This table reports the out-of-sample performance of sophisticated forecast models estimated using recursively expanding windows or Pesaran and Timmermann (2007)'s average windows forecasting (AveW) method. Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. These sophisticated models are the standard multivariate regression model (Kitchen sink), the weighted-average least squares (WALS) introduced by Magnus et al. (2010) , Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. For comparison, we also report the results of the simple equal-weighted combination of individual predictors from Rapach et al. (2010), called Pool-AVG. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the forecast model given in the first column relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. The out-of-sample evaluation period is 1957:01–2016:12.

| Model | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ | Time consumed (s) |
|---|---|---|---|---|---|
| Pool-AVG | 0.42 | 1.90** | 0.11 | 1.11 | 20 |
| **Panel A: Estimated using recursively expanding windows** | | | | | |
| Kitchen sink | -7.85 | 0.36 | -8.57 | -6.23 | 7 |
| WALS | -4.14 | 0.21 | -4.79 | -2.67 | 82 |
| MMA | -3.67 | 0.17 | -4.05 | -2.81 | 195 |
| JMA | -0.72 | -0.16 | -0.96 | -0.17 | 329 |
| Elastic net | 0.04 | 0.57 | 0.27 | -0.50 | 554 |
| LASSO | -0.24 | 0.02 | -0.31 | -0.07 | 614 |
| BMA | -2.03 | -1.34 | -1.34 | -3.60 | 2,391 |
| **Panel B: Estimated using average windows forecasting method** | | | | | |
| Kitchen sink | -1.97 | 1.80** | -1.57 | -2.86 | 34 |
| WALS | 0.09 | 1.98** | 0.29 | -0.35 | 612 |
| MMA | -0.49 | 1.72** | -0.47 | -0.53 | 958 |
| JMA | 0.33 | 1.66** | 0.39 | 0.20 | 1,587 |
| Elastic net | 0.47 | 1.66* | 0.35 | 0.73 | 3,437 |
| LASSO | 0.57 | 1.88** | 0.38 | 1.85 | 3,264 |
| BMA | -0.92 | 1.65** | -1.50 | 0.41 | 20,550 |

**Table 3**
**Out-of-sample forecasting results of the simple combining method, 1957:01–2016:12**

This table reports the out-of-sample performance of the equal-weighted combinations of the historical average (HA) forecast with different sophisticated models, including the standard multivariate regression model (HA+Kitchen sink), Magnus et al. (2010)'s weighted-average least squares (HA+WALS), Hansen (2007, 2008)'s Mallows model averaging (HA+MMA), Hansen and Racine (2012)'s jackknife model averaging (HA+JMA), LASSO and Elastic net with 5-fold cross-validation (HA+LASSO and HA+Elastic net), and BMA with diffuse prior (HA+BMA). For comparison, we also report the results of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG. All these models including Pool-AVG are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. $R^2_{OS,exp}$ ($R^2_{OS,rec}$) statistics are calculated over NBER-dated business-cycle expansions (recessions). The out-of-sample evaluation period is 1957:01–2016:12.

| Combinations | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ |
|---|---|---|---|---|
| Pool-AVG | 0.45 | 1.99** | 0.06 | 1.31 |
| HA + Kitchen sink | 1.10 | 1.96** | 1.21 | 0.86 |
| HA + WALS | 1.22 | 1.93** | 1.31 | 1.04 |
| HA + MMA | 0.95 | 1.74** | 0.86 | 1.15 |
| HA + JMA | 0.94 | 1.69** | 0.87 | 1.11 |
| HA + Elastic net | 0.84 | 1.66** | 0.69 | 1.18 |
| HA + LASSO | 0.99 | 1.72** | 0.93 | 1.11 |
| HA + BMA | 0.12 | 0.91 | -0.24 | 0.93 |

**Table 4**
**Asset allocation results of the simple combining method, 1957:01–2016:12**

This table reports asset allocation results of the equal-weighted combinations of the historical average (HA) forecast with different sophisticated models, including the standard multivariate regression model (HA+Kitchen sink), Magnus et al. (2010)'s weighted-average least squares (HA+WALS) , Hansen (2007, 2008)'s Mallows model averaging (HA+MMA), Hansen and Racine (2012)'s jackknife model averaging (HA+JMA), LASSO and Elastic net with 5-fold cross-validation (HA+LASSO and HA+Elastic net), and BMA with diffuse prior (HA+BMA). For comparison, we also report the results of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG. All these models including Pool-AVG are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. The utility gain $\triangle(ann\%)$ is the annualized certainty equivalent return gain for the investor with mean-variance preferences and risk aversion coefficient $\gamma = 3$. The monthly Sharpe ratio is the mean portfolio return in excess of the risk-free rate divided by its standard deviation. The out-of-sample evaluation period is 1957:01–2016:12.

| Model | No transaction cost | | 50bps transaction cost | |
|---|---|---|---|---|
| | $\triangle(ann\%)$ | Sharpe ratio | $\triangle(ann\%)$ | Sharpe ratio |
| Pool-AVG | 0.87 | 0.12 | 0.31 | 0.11 |
| HA + Kitchen sink | 2.34 | 0.14 | 0.54 | 0.11 |
| HA + WALS | 2.46 | 0.14 | 0.91 | 0.12 |
| HA + MMA | 2.22 | 0.14 | 0.73 | 0.11 |
| HA + JMA | 1.72 | 0.13 | 0.44 | 0.11 |
| HA + Elastic net | 1.94 | 0.14 | 0.82 | 0.12 |
| HA + LASSO | 1.64 | 0.13 | 0.50 | 0.11 |
| HA + BMA | 1.01 | 0.12 | 0.07 | 0.10 |

**Table 5** 涵盖性检验

**Forecast encompassing test for the simple combining method, MHLN statistic p-values, 1957:01–2016:12**

The table reports p-values of the Harvey et al. (1998) MHLN statistic for the historical average (HA) forecast and different equal-weighted combinations, including the combination of individual predictors (Pool-AVG), the combinations of HA with different sophisticated models. These sophisticated models are the standard multivariate regression model (Kitchen sink), the weighted-average least squares (WALS) introduced by Magnus et al. (2010) , Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. All these models are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). The MHLN statistic corresponds to a one-sided (upper-tail) test of the null hypothesis that the forecast under the scheme given in the column heading encompasses the forecast under the scheme given in the row heading against the alternative hypothesis that the forecast under the scheme given in the column heading does not encompass the forecast under the scheme given in the row heading. 0.00 indicates less than 0.005. The MHLN statistic is computed for the entire 1957:01–2016:12 forecast evaluation period.

| | HA | Pool-AVG | HA+Kitchen sink | HA+WALS | HA+MMA | HA+JMA | HA+ElasticNet | HA+LASSO | HA+BMA |
|---|---|---|---|---|---|---|---|---|---|
| HA | | 0.89 | 0.27 | 0.52 | 0.42 | 0.57 | 0.61 | 0.62 | 0.28 |
| Pool-AVG | 0.02 | | 0.24 | 0.45 | 0.35 | 0.47 | 0.52 | 0.55 | 0.14 |
| HA+Kitchen sink | 0.03 | 0.06 | | 0.47 | 0.23 | 0.20 | 0.14 | 0.18 | 0.05 |
| HA+WALS | 0.03 | 0.06 | 0.27 | | 0.04 | 0.15 | 0.12 | 0.15 | 0.03 |
| HA+MMA | 0.04 | 0.10 | 0.48 | 0.92 | | 0.39 | 0.24 | 0.35 | 0.04 |
| HA+JMA | 0.05 | 0.11 | 0.35 | 0.65 | 0.40 | | 0.23 | 0.45 | 0.02 |
| HA+ElasticNet | 0.05 | 0.14 | 0.32 | 0.59 | 0.40 | 0.53 | | 0.72 | 0.0 |
| HA+LASSO | 0.04 | 0.12 | 0.27 | 0.52 | 0.28 | 0.29 | 0.17 | | 0.02 |
| HA+BMA | 0.18 | 0.50 | 0.56 | 0.88 | 0.81 | 0.93 | 0.95 | 0.95 | |

**Table 6**
**Forecasting extreme market movements with the simple combining method**

This table reports the $R^2_{OS}$ values of the equal-weighted combinations of the historical average (HA) forecast with the standard multivariate regression model (HA+Kitchen sink) or Magnus et al. (2010)'s weighted-average least squares (HA+WALS), for extreme (downturn) periods of monthly returns $r_t$ that are normalized with mean 0 and standard deviation of 1. For comparison, the results of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG, are also reported. All these models are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively, based on p-values for the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. The out-of-sample evaluation period is 1957:01–2016:12.

| Periods | Observations | HA + Kitchen sink | HA + WALS | Pool-AVG |
|---|---|---|---|---|
| Overall | 720 / 720 | 1.10** | 1.22** | 0.45** |
| **Panel A: extreme periods of normalized returns $r_t$** | | | | |
| $\|r_t\| \geq 0.5$ | 419 / 720 | 1.82** | 1.65** | 0.57** |
| $\|r_t\| \geq 1.0$ | 189 / 720 | 2.86** | 2.29** | 0.85** |
| $\|r_t\| \geq 1.5$ | 91 / 720 | 2.94* | 2.37* | 0.94* |
| $\|r_t\| \geq 2.0$ | 40 / 720 | 3.94* | 3.19* | 1.05* |
| **Panel B: extreme downturn periods of normalized returns $r_t$** | | | | |
| $r_t \leq -0.5$ | 199 / 720 | 5.17*** | 3.83*** | 2.10*** |
| $r_t \leq -1.0$ | 100 / 720 | 6.01*** | 4.44*** | 2.03*** |
| $r_t \leq -1.5$ | 54 / 720 | 6.38*** | 4.79** | 1.81*** |
| $r_t \leq -2.0$ | 25 / 720 | 6.84** | 5.32** | 1.49*** |

**Table 7**
**Forecasting characteristic portfolios with the simple combining method**

This table reports the performance of the equal-weighted combination of the historical average forecast with the standard multivariate regression model Kitchen sink (HA + Kitchen sink), for the 10 size characteristic portfolios. For comparison, the results of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG, are also reported. Both Kitchen sink and Pool-AVG are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively, based on p-values for the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. The utility gain $\triangle(ann\%)$ is the annualized certainty equivalent return gain for the investor with mean-variance preferences and risk aversion coefficient $\gamma = 3$. The monthly Sharpe ratio is the mean portfolio return in excess of the risk-free rate divided by its standard deviation. Size portfolio returns are value-weighted and available from Kenneth French's data library. The out-of-sample evaluation period is 1957:01–2016:12.

| **Size portfolios** | **HA + Kitchen sink** | | | **Pool-AVG** | | |
|---|---|---|---|---|---|---|
| | $R^2_{OS}(\%)$ | $\triangle(ann\%)$ | Sharpe ratio | $R^2_{OS}(\%)$ | $\triangle(ann\%)$ | Sharpe ratio |
| Small | 2.58 | 5.53 | 0.24 | 0.67 | 1.31 | 0.19 |
| 2 | 0.94 | 7.18 | 0.22 | 0.46 | 4.70 | 0.18 |
| 3 | 0.99 | 7.72 | 0.24 | 0.49 | 5.16 | 0.20 |
| 4 | 0.94 | 7.23 | 0.23 | 0.60 | 5.19 | 0.20 |
| 5 | 0.81 | 7.27 | 0.23 | 0.58 | 5.05 | 0.20 |
| 6 | 0.85 | 6.61 | 0.24 | 0.61 | 5.14 | 0.21 |
| 7 | 0.85 | 7.33 | 0.25 | 0.51 | 5.02 | 0.21 |
| 8 | 0.60 | 6.52 | 0.24 | 0.37 | 4.80 | 0.21 |
| 9 | 0.66 | 6.04 | 0.24 | 0.38 | 4.51 | 0.22 |
| Large | 0.67 | 5.21 | 0.22 | 0.32 | 3.83 | 0.20 |

**Table 8**
**Forecasting macro variables with the simple combining method**

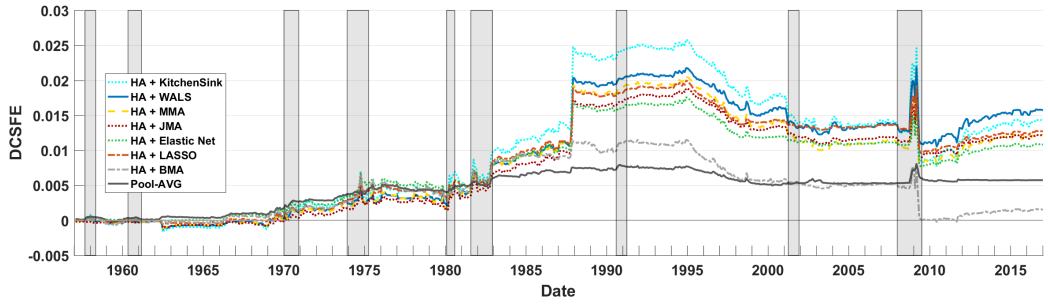
This table reports the $R^2_{OS}$ values of the equal-weighted combination of the historical average (HA) forecast with the standard multivariate regression model Kitchen sink, for a list of macroeconomic condition variables. The Kitchen sink model is estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $Y_{t+1}$ is the macroeconomic condition for next month, $Y_{t+12}$ is the macroeconomic condition for next year, $\Delta Y_{t+1} = Y_{t+1} - Y_{(t+1)-12}$ is the year change in macroeconomic condition for the next month, and $\Delta Y_{t+12} = Y_{t+12} - Y_t$ is the year change in macroeconomic condition for the next year. These macro condition variables are Chicago Fed National Activity Index (CFNAI), Smoothed U.S. Recession Probabilities (SRP), Industrial Production Growth (IPG), Jurado et al. (2015)'s Macroeconomic Uncertainty Index (MU), Output Gap (Gap), Lettau and Ludvigson (2001a,b)'s Cay that is a cointegrating residual between log consumption, log asset (nonhuman) wealth, and log labor income, Civilian Unemployment Rate (UNRATE). $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively, based on p-values for the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. The out-of-sample evaluation period given in column two depends on the data sample available.

| Macro Variable | Evaluation Period | $Y_{t+1}$ | $\triangle Y_{t+1}$ | $Y_{t+12}$ | $\triangle Y_{t+12}$ |
|---|---|---|---|---|---|
| CFNAI | 2000:01-2016:12 | 34.78*** | 13.48*** | 10.10*** | 11.55*** |
| SRP | 2000:01-2016:12 | 40.56*** | 10.80*** | 13.80*** | 19.50*** |
| IPG | 1957:01-2016:12 | 60.62*** | 45.87*** | 60.85*** | 21.57*** |
| MU | 1995:01-2016:12 | 19.03*** | 15.51*** | 35.79*** | 15.98*** |
| GAP | 1980:01-2016:12 | 52.61*** | 37.74*** | 34.00*** | 31.53*** |
| Cay | 1985:01-2016:12 | 23.14*** | 9.05*** | 14.84*** | 7.51*** |
| UNRATE | 1980:01-2016:12 | 50.36*** | 49.25*** | 43.55*** | 31.49*** |

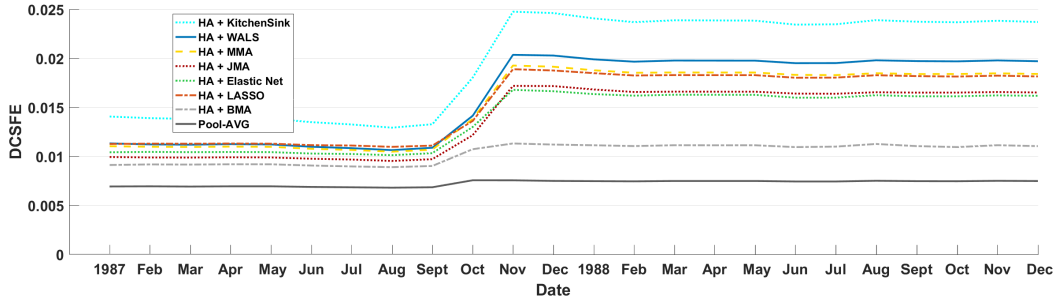
**Table 9**
**Robustness check of the simple combining method based on factors extracted from the FRED-MD dataset**

This table reports the $R^2_{OS}$ values of the equal-weighted combinations of the historical average (HA) forecast with different sophisticated models, based on factors extracted from FRED-MD, a macroeconomic database of 134 monthly U.S. indicators (Mccracken and Ng, 2015). The data spans from 1959:01 through 2016:12. These sophisticated models are the standard multivariate regression model (Kitchen sink), the weighted-average least squares (WALS) introduced by Magnus et al. (2010) , Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. For comparison, the results of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG, are also reported. All these models are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the forecast model given in the first column relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. The out-of-sample evaluation period is 1980:01–2016:12.

| Model | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ |
|---|---|---|---|---|
| Pool-AVG | 0.63 | 1.82** | 0.37 | 1.34 |
| HA + Kitchen sink | 0.93 | 1.69** | -0.32 | 4.38 |
| HA + WALS | 0.97 | 1.71** | 0.04 | 3.53 |
| HA + MMA | 0.92 | 1.70** | 0.10 | 3.16 |
| HA + JMA | 0.82 | 1.60* | 0.17 | 2.61 |
| HA + Elastic net | 0.94 | 1.82** | 0.41 | 2.40 |
| HA + LASSO | 0.99 | 1.86** | 0.38 | 2.67 |
| HA + BMA | 0.98 | 1.83** | 0.28 | 2.88 |




(a) 1957:01–2016:12




(b) 1987:01–1988:12

**Fig. 1. Forecast accuracy of the simple combining method over time**

This figure shows the relative performance of the equal-weighted combinations of the historical average (HA) forecast with different sophisticated models, compared to the historical average benchmark model, in terms of the DCSFE, which is defined as the difference in the cumulative squared forecast error of the historical average benchmark model and the given model. The larger the DCSFE value, the better the model's performance. These sophisticated models are the standard multivariate regression model (Kitchen sink), the weighted-average least squares (WALS) introduced by Magnus et al. (2010) , Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. For comparison, we also report the relative performance of the equal-weighted combining method for individual predictors from Rapach et al. (2010), called Pool-AVG. All these models are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007). Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. The vertical bars in figure (a) correspond to the NBER-dated recessions. Figure (b) depicts the performance of different models during the sub period of 1987:01–1988:12. The out-of-sample evaluation period is 1957:01–2016:12.

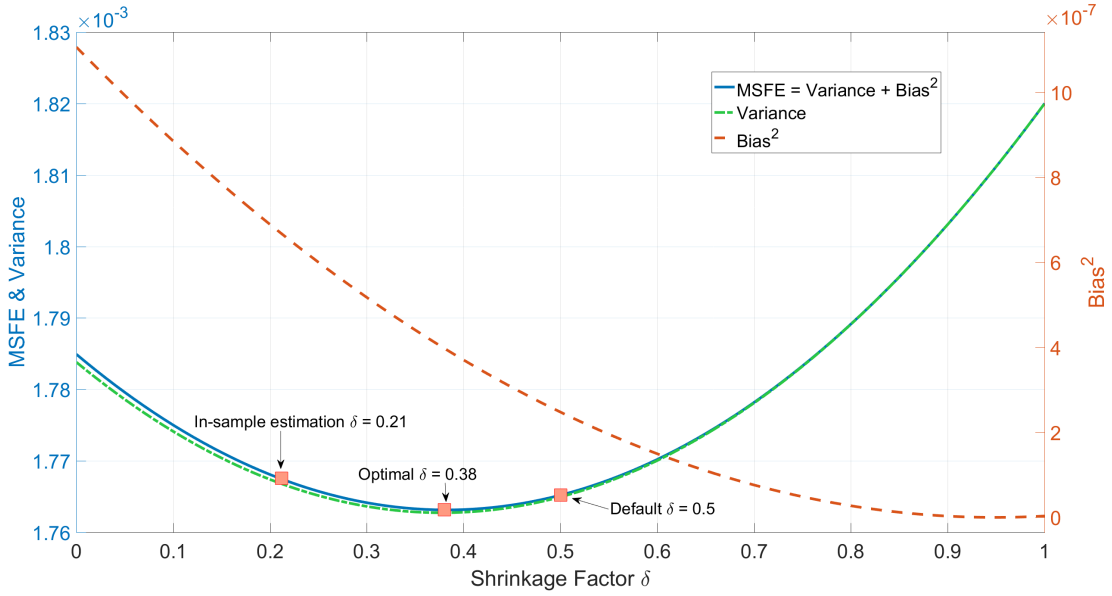



**Fig. 2. Effects of shrinkage**

This figure illustrates the bias-variance decomposition of the MSFE of the equal-weighted combination of the historical average with the standard multivariate regression model, against the shrinkage factor $\delta \in \{0.00, 0.01, 0.02, ..., 0.99, 1.00\}$. The bias-variance decomposition of MSFE (blue solid line) is defined as $MSFE = Bias^2 + Variance$, where $Bias^2$ and $Variance$ are the mean squared forecast bias (red dash line) and mean forecast error variance (green dash-dot line) of the equal-weighted combination, respectively. The out-of-sample forecast evaluation period is 1957:01–2016:12.

# Internet Appendix

**Table A.1**
**Robust analysis of the averaging windows forecasting method**

This table reports out-of-sample $R^2_{OS}$ values of different sophisticated forecast models estimated using Pesaran and Timmermann (2007)'s averaging windows (AveW) forecasting method with different numbers of estimation windows $m \in \{10, 20, 30, 40, 50, 100\}$. These sophisticated models are the standard multivariate regression model (Kitchen sink), weighted-average least squares (WALS) introduced by Magnus et al. (2010), Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the forecast model given in the first column relative to the historical average benchmark forecast. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively, based on p-values for the Clark and West (2007) MSFE-adjusted statistic. The out-of-sample evaluation period is 1957:01–2016:12.

| Model \ m | 10 | 20 | 30 | 40 | 50 | 100 |
|---|---|---|---|---|---|---|
| Kitchen sink | $-1.97^{**}$ | $-1.83^{**}$ | $-1.81^{**}$ | $-1.89^{**}$ | $-1.85^{**}$ | $-1.83^{**}$ |
| WALS | $0.09^{**}$ | $0.17^{**}$ | $0.16^{**}$ | $0.08^{**}$ | $0.11^{**}$ | $0.13^{**}$ |
| MMA | $-0.49^{**}$ | $-0.43^{**}$ | $-0.40^{**}$ | $-0.46^{**}$ | $-0.38^{**}$ | $-0.43^{**}$ |
| JMA | $0.33^{**}$ | $0.24^{**}$ | $0.26^{**}$ | $0.22^{**}$ | $0.29^{**}$ | $0.25^{**}$ |
| Elastic Net | $0.47^{*}$ | $0.64^{**}$ | $0.31^{*}$ | $0.39^{*}$ | $0.17^{*}$ | $0.33^{*}$ |
| LASSO | $0.57^{**}$ | $0.54^{*}$ | $0.65^{**}$ | $0.58^{**}$ | $0.45^{**}$ | $0.36^{*}$ |
| BMA | $-0.92$ | $-0.91$ | $-0.88$ | $-0.95$ | $-0.91$ | $-0.92$ |

**Table A.2**
**Out-of-sample forecasting results of the simple combining method estimated traditionally, 1957:01–2016:12**

This table reports the out-of-sample performance of the equal-weighted combinations of the historical average (HA) forecast with different sophisticated models, including the standard multivariate regression model (HA+Kitchen sink), Magnus et al. (2010)'s weighted-average least squares (HA+WALS), Hansen (2007, 2008)'s Mallows model averaging (HA+MMA), Hansen and Racine (2012)'s jackknife model averaging (HA+JMA), LASSO and Elastic net with 5-fold cross-validation (HA+LASSO and HA+Elastic net), and BMA with diffuse prior (HA+BMA). All these models are estimated using recursively expanding windows. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. $R^2_{OS,exp}$ ($R^2_{OS,rec}$) statistics are calculated over NBER-dated business-cycle expansions (recessions). The out-of-sample evaluation period is 1957:01–2016:12.

| Combinations | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ |
|---|---|---|---|---|
| HA + Kitchen sink | -1.70 | 0.36 | -2.26 | -0.42 |
| HA + WALS | -0.92 | 0.21 | -1.37 | 0.09 |
| HA + MMA | -0.84 | 0.17 | -1.15 | -0.12 |
| HA + JMA | -0.21 | −0.16 | -0.38 | 0.16 |
| HA + Elastic Net | 0.04 | 0.57 | 0.16 | -0.23 |
| HA + LASSO | -0.09 | −0.78 | -0.12 | -0.03 |
| HA + BMA | -0.80 | −1.34 | -0.47 | -1.56 |

**Table A.3**
**Out-of-sample forecasting results of the performance-based combining method**

This table reports the out-of-sample performance of the performance-based combinations of the historical average (HA) forecast with different sophisticated models, including the standard multivariate regression model (HA+Kitchen sink), Magnus et al. (2010)'s weighted-average least squares (HA+WALS), Hansen (2007, 2008)'s Mallows model averaging (HA+MMA), Hansen and Racine (2012)'s jackknife model averaging (HA+JMA), LASSO and Elastic net with 5-fold cross-validation (HA+LASSO and HA+Elastic net), and BMA with diffuse prior (HA+BMA). Following Rapach et al. (2010), the combining weights are computed based on the forecasting performance (in terms of the MSFE) of individual models over a holdout out-of-sample period 120 months. For comparison, we also report the results of the performance-based combining method for individual predictors from Rapach et al. (2010), called Pool-DMSFE. When these models are estimated using the average windows (AveW) forecasting method introduced by Pesaran and Timmermann (2007), following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the given forecast model relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. $R^2_{OS,exp}$ ($R^2_{OS,rec}$) statistics are calculated over NBER-dated business-cycle expansions (recessions). The out-of-sample evaluation period is 1957:01–2016:12.

| Combinations | $R^2_{OS}(\%)$ | CW-test | No transaction cost | | 50bps transaction cost | |
|---|---|---|---|---|---|---|
| | | | $\triangle(ann\%)$ | Sharpe ratio | $\triangle(ann\%)$ | Sharpe ratio |
| **Panel A: Estimated using recursively expanding windows** | | | | | | |
| Pool-DMSFE | 0.45 | 1.98** | 0.78 | 0.12 | .41 | 0.11 |
| HA + Kitchen sink | -1.74 | 0.23 | 0.41 | 0.11 | -1.17 | 0.08 |
| HA + WALS | -0.99 | 0.12 | 0.88 | 0.12 | -0.40 | 0.09 |
| HA + MMA | -0.91 | 0.07 | 0.82 | 0.11 | -0.42 | 0.09 |
| HA + JMA | -0.24 | −0.19 | 0.37 | 0.11 | -0.19 | 0.10 |
| HA + Elastic net | 0.04 | 0.57 | -0.09 | 0.10 | -0.63 | 0.10 |
| HA + LASSO | -0.09 | −0.78 | -0.11 | 0.10 | -0.67 | 0.10 |
| HA + BMA | -0.78 | −1.32 | -0.64 | 0.09 | -0.98 | 0.09 |
| **Panel B: Estimated using average windows forecasting method** | | | | | | |
| Pool-DMSFE | 0.46 | 1.90** | 1.00 | 0.12 | 0.40 | 0.11 |
| HA + Kitchen sink | 0.98 | 1.86** | 2.42 | 0.14 | 0.60 | 0.11 |
| HA + WALS | 1.15 | 1.85** | 2.55 | 0.15 | 1.00 | 0.12 |
| HA + MMA | 0.87 | 1.67** | 2.25 | 0.14 | 0.74 | 0.11 |
| HA + JMA | 0.92 | 1.65** | 1.81 | 0.13 | 0.53 | 0.11 |
| HA + Elastic net | 0.84 | 1.64* | 1.99 | 0.14 | 0.85 | 0.12 |
| HA + LASSO | 0.99 | 1.70** | 1.73 | 0.13 | 0.58 | 0.11 |
| HA + BMA | 0.12 | 0.91 | 1.16 | 0.12 | 0.21 | 0.10 |

**Table A.4**
**Robustness check of sophisticated models estimated traditionally**

This table reports the out-of-performance of different sophisticated models, based on factors extracted from FRED-MD, a macroeconomic database of 134 monthly U.S. indicators (Mccracken and Ng, 2015). The data spans from 1959:01 through 2016:12. These sophisticated models are the standard multivariate regression model (Kitchen sink), the weighted-average least squares (WALS) introduced by Magnus et al. (2010) , Mallows model averaging (MMA) proposed by Hansen (2007, 2008), jackknife model averaging (JMA) proposed by Hansen and Racine (2012), LASSO and Elastic net with 5-fold cross-validation, and BMA with diffuse prior. For comparison, we also report the results of the simple equal-weighted combination of individual predictors from Rapach et al. (2010), called Pool-AVG. All these models are estimated using recursively expanding windows or Pesaran and Timmermann (2007)'s average windows forecasting (AveW) method. Following Pesaran and Pick (2011), we set the number of estimation windows of AveW $m = 10$. $R^2_{OS}$ measures the percent reduction in mean squared forecast error (MSFE) for the forecast model given in the first column relative to the historical average benchmark forecast. CW-test is the Clark and West (2007) MSFE-adjusted statistic for testing the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively. The out-of-sample evaluation period is 1980:01–2016:12.

| Model | $R^2_{OS}(\%)$ | CW-test | $R^2_{OS,exp}(\%)$ | $R^2_{OS,rec}(\%)$ |
|---|---|---|---|---|
| **Panel A: Estimated using recursively expanding windows** | | | | |
| Pool-AVG | 0.55 | 1.83** | 0.32 | 1.18 |
| Kitchen sink | -2.46 | 1.32* | -3.14 | -0.59 |
| WALS | -0.99 | 1.22 | -1.61 | 0.69 |
| MMA | -0.13 | 1.87** | -1.58 | 3.84 |
| JMA | -0.02 | 1.76** | -1.27 | 3.43 |
| Elastic net | 0.05 | 1.92** | -1.38 | 3.99 |
| LASSO | 0.14 | 1.79** | -0.91 | 3.01 |
| BMA | 0.11 | 1.99** | -1.24 | 3.84 |
| **Panel B: Estimated using average windows forecasting method** | | | | |
| Pool-AVG | 0.63 | 1.82** | 0.37 | 1.34 |
| Kitchen sink | -1.18 | 1.69** | -3.08 | 4.06 |
| WALS | 0.22 | 1.71** | -1.31 | 4.45 |
| MMA | 0.12 | 1.70** | -1.18 | 3.73 |
| JMA | 0.22 | 1.60* | -0.84 | 3.13 |
| Elastic net | 0.39 | 1.70** | -0.49 | 2.81 |
| LASSO | 0.65 | 1.81** | -0.25 | 3.11 |
| BMA | 0.67 | 1.83** | -0.47 | 3.79 |