Improving the Simple Average Combined Forecast via Factor-Adjusted Regularization

Saerom Lee* and Tae-Hwy Lee†

March 22, 2026

Abstract

This paper addresses the forecast combination puzzle—the empirical observation that the simple average forecast combination often outperforms complex weighting schemes—by employing a factor-adjusted regularization framework. In this framework, the simple average is treated as a common factor. We identify and incorporate idiosyncratic components, defined as the deviations of individual forecasts from the simple average. Our approach effectively manages the high correlation among forecasts by focusing on idiosyncratic components that enhance predictive content beyond what is captured by the simple average. Empirical applications in macroeconomic forecasting show that this factor-adjusted approach yields significant accuracy gains over the simple average.

Keywords: simple average, common factors, idiosyncratic components, forecast combination puzzle, encompassing, high dimension, sparsity

JEL classification: C22, C32

*Department of Economics, The University of Toledo, Toledo, OH 43606. E-mail: saeromlee3390@gmail.com.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu.

1 Introduction

Combining multiple forecasts (Bates and Granger (1969)) has been a widely adopted strategy to produce forecasts that outperform those of any single model. The empirical success of this approach is often attributed to the presence of structural breaks and other instabilities (Rossi and Sekhposyan (2013) and Rossi (2021)).

Despite this empirical success, the widespread use of forecast combination still faces several challenges, such as selecting which forecasts to include and determining the optimal method of combination. A particularly notable empirical issue is the "forecast combination puzzle". Coined by Stock and Watson (2004) and also noted by Chan et al. (1999), this puzzle refers to the recurring empirical finding that a simple average of forecasts using equal weights often outperforms more sophisticated combinations based on estimated optimal weights. In particular, as Smith and Wallis (2009) note, "When the number of competing forecasts is large, so that under equal weighting each has a very small weight, the simple average can gain in efficiency by trading off a small bias against a larger estimation variance". This implies that the simple average combined forecast performs relatively well when many forecasts are available.

The existing literature provides substantial empirical evidence on this puzzle. For example, Clemen and Winkler (1986) find that the simple average performs well when combining forecasts from four econometric models (Wharton, Chase, Data Resources Inc., and the Bureau of Economic Analysis) to predict quarterly real and nominal GNP growth. Similarly, Genre et al. (2013) use data from the European Central Bank Survey of Professional Forecasters and find that only a few combination methods outperform the simple average in forecasting the unemployment rate and GDP growth. Green and Armstrong (2015) review 32 papers comparing complex and simple forecast combination methods, concluding that, in most cases, complexity harms forecast accuracy. Frazier et al. (2023) attribute the forecast combination puzzle to the low power and lack of size control in existing tests, which tend to favor the simple average. Wang et al. (2023) provide a review of the extensive literature on forecast combinations, specifically highlighting the persistence of the forecast combination puzzle.

However, it is unlikely that all individual forecasts possess the same level of predictive accuracy. In light of this, we propose a factor-adjusted combined forecast designed to outperform the simple average by selectively integrating signals from individual forecasts that capture heterogeneity. In our approach, the simple average is treated as the primary common factor shared across all individual forecasts. Starting with the simple average as an initial learner, we aim to improve performance by incorporating idiosyncratic components only when they reduce the forecast error loss. In other words, the combined forecast is updated only when it becomes a stronger learner.

We demonstrate the empirical performance of our method using monthly U.S. macroeconomic time series from the Federal Reserve Economic Data (FRED). The results show that selectively incorporating idiosyncratic components yields significant predictive gains, demonstrating their crucial role in enhancing the predictive power of the simple average.

The remainder of the paper is organized as follows. Section 2 presents a brief literature review on the forecast combination puzzle. Section 3 introduces our new combined forecast focusing on the role of selected idiosyncratic components in improving the simple average. Section 4 presents the empirical applications in macroeconomic forecasting. Section 5 concludes.

2 Forecast Combination Puzzle

A combined forecast of N individual forecasts, $f_{1,t+h}, \ldots, f_{N,t+h}$, for the target variable y_{t+h}, is defined as follows:

$$f_{c,t+h} = \sum_{i=1}^{N} \beta_i f_{i,t+h}, \quad t = 1, \ldots, T, \tag{1}$$

where h denotes the forecast horizon at time t. The coefficients β_i represent the combining weights, which sum to one: $\sum_{i=1}^{N} \beta_i = 1$. Bates and Granger (1969) show that the mean squared forecast error (MSFE) of the combined forecast is minimized with the optimal weight vector $\boldsymbol{\beta}^* = \frac{\Sigma^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\iota}}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)'$, Σ is the covariance matrix of the N forecast errors ($u_{i,t+h} = y_{t+h} - f_{i,t+h}$), and $\boldsymbol{\iota}$ is a column vector of ones. Granger and Ramanathan (1984) (hereafter GR) use the ordinary least squares (OLS) regression to estimate the combining weights β_i.

However, the population covariance matrix of forecast errors, Σ, is unknown and must be estimated. In practice, finite-sample estimation errors can significantly reduce the benefits of combining forecasts (Smith and Wallis (2009), Claeskens et al. (2016), and Chan and Pauwels (2018)). This is particularly true in high-dimensional settings, where the number of forecasts N is large relative to the number of observations T. The instability in the estimated optimal weights often leads to poor out-of-sample performance of the combined forecast. A similar issue arises with the regression-based approach, which encounters difficulties when N is close to or exceeds T (Smith and Wallis (2009)). For these reasons, the simple average combined forecast maintains its appeal, offering a robust alternative that requires no parameter estimation.

Furthermore, numerous studies have proposed explanations, particularly through the lens of shrinkage, for the robust performance of the simple average. This concept involves adjusting estimated weights closer to equal values; for instance, Diebold and Pauly (1990) applied Bayesian shrinkage to pull estimated weights toward equality. More recently, this idea has been extended to shrinking the weights of a subset of forecasts. Since the OLS regression becomes computationally infeasible in high-dimensional settings, researchers have increasingly adopted penalization and regularization methods. These methods simultaneously select a subset of forecasts for combination and shrink their weights toward equal weights. For example, Matsypura et al. (2018) propose an integer programming method for optimal subset selection, while Liu et al. (2024) develop a double shrinkage method using weighted least squares. Elliott and Liao (2025) provide further evidence that simply discarding some forecasts and averaging over a selected subset can significantly improve forecast accuracy.

Among the various combination methods developed in the literature, we focus more specifically on the approach of Diebold and Shin (2019) (hereafter DS). This is because their method has a close relationship, via a reparametrization, with the one we introduce in Section 3. Diebold and Shin (2019) present a method based on the Least Absolute Shrinkage and Selection Operator (Lasso). This method shrinks the combining weights (β_i) toward the equal weights ($\frac{1}{N}$) and is referred to as the "egalitarian Lasso". The combined forecast $f_{c,t+h}$ in Equation (1) is reformulated by adding and subtracting the simple average as follows:

$$
\begin{aligned}
f_{c,t+h} &= \frac{1}{N}\sum_{i=1}^{N} f_{i,t+h} + \sum_{i=1}^{N}\left(\beta_i - \frac{1}{N}\right) f_{i,t+h} \\
&= \bar{f}_{t+h} + \sum_{i=1}^{N} \delta_i f_{i,t+h},
\end{aligned}
\tag{2}
$$

where $\bar{f}_{t+h} \equiv \frac{1}{N}\sum_{i=1}^{N} f_{i,t+h}$ is the simple average combined forecast, and $\delta_i \equiv \beta_i - \frac{1}{N}$.

By re-expressing Equation (2) in terms of forecast errors, we have the following regression equation:

$$
\bar{u}_{t+h} = \sum_{i=1}^{N} \delta_i f_{i,t+h} + u_{c,t+h},
\tag{3}
$$

where $\bar{u}_{t+h} \equiv y_{t+h} - \bar{f}_{t+h}$ is the forecast error of the simple average and $u_{c,t+h} \equiv y_{t+h} - f_{c,t+h}$ is the forecast error of the combined method. The coefficients δ_i can then be estimated by regressing $\bar{u}_{t+h}$ on the individual forecasts $f_{i,t+h}$.

In high-dimensional settings, where the number of forecasts N is large, large variance in estimation can be mitigated by applying regularization techniques like Lasso, Ridge, and L_2-Boosting to the coefficients δ_i. These techniques impose a penalty on the magnitude of the coefficients, effectively inducing shrinkage by driving some δ_i toward zero. This regularization simultaneously addresses high-dimensionality and operationalizes the shrinkage principle: as the δ_i coefficients approach zero, the combined forecast $f_{c,t+h}$ is explicitly shrunk toward the simple average $\bar{f}_{t+h}$.

However, the success of the Lasso-based approach rests on the assumption that the true coefficients are sparse (i.e., many coefficients δ_i are zero, implying many individual forecasts are irrelevant). This assumption is violated when there are additional common factors among individual forecasts beyond the simple average. In such cases, the coefficients δ_i may not be sparse. See the Online Supplementary Appendix for a detailed discussion.

Although Diebold and Shin (2019) include an initial selection step, its purpose is primarily to screen out redundant forecasts rather than to ensure strict sparsity. They discard some forecasts and then shrink the remaining forecasts toward equal weights, which is referred to as the "partially egalitarian Lasso". Furthermore, the Lasso's ability to consistently perform model selection (i.e., correctly identifying non-zero δ_i) is known to depend on the "irrepresentable condition"(Zhao and Yu (2006)). This condition, which is (almost) necessary and sufficient for consistent selection, requires that irrelevant predictors be "irrepresentable" by relevant ones. This condition would be violated in Equation (3) for DS, as the individual forecasts $f_{i,t+h}$ are highly correlated in the presence of omitted relevant factors.

To address the aforementioned issue, we reformulate the DS model (Equations (2) and (3)) into our proposed model (Equations (5) and (6) in Section 3; see also Equations (A.2) and (A.3) in the Online Supplementary Appendix). Both pairs of equations are derived from the decomposition of individual forecasts into common factors and idiosyncratic components. This framework follows the Factor Adjusted Regularized Model (FARM) approach of Fan et al. (2020), Fan et al. (2023), and Fan et al. (2024), which reconciles a dense factor structure with sparse idiosyncratic components.

3 Improving the Simple Average with Idiosyncratic Components

In this section, we introduce a new combined forecast designed to improve upon the simple average combined forecast. Our approach augments the simple average by incorporating selected idiosyncratic components, ensuring the combined forecast $f_{c,t+h}$ encompasses the simple average forecast $\bar{f}_{t+h}$. A forecast is said to encompass another if it yields a smaller forecast error loss (Harvey et al. (1998)).

We postulate that the N forecasts follow a factor structure, given their high correlation due to shared predictors and their common target variable. Previous studies (Figlewski and Urich (1983), Chan et al. (1999), and Poncela et al. (2011)) have also considered factor structures in forecast combination. In contrast to their approaches, our proposed method accounts not only for the common factors but also for selected idiosyncratic components. Beyhum and Striaukas (2024) suggest that sparse idiosyncratic components may play an additional role alongside factors in macroeconomic and finance. Moreover, idiosyncratic components capture signals that can improve predictive performance when over- and under-predictions have asymmetric consequences for the target variable (Elliott et al. (2005) and Elliott et al. (2008)).

We consider the case where the simple average is treated as the sole common factor among individual forecasts. We decompose each individual forecast $f_{i,t+h}$ into this common factor and an idiosyncratic component, as follows:

$$f_{i,t+h} = \bar{f}_{t+h} + d_{i,t+h}, \tag{4}$$

where $d_{i,t+h}$ denotes the idiosyncratic component. These idiosyncratic components are weakly correlated (or uncorrelated) after accounting for the common factor, and they may contain forecast-specific predictive signals beyond noise. For instance, in forecasting unemployment, the idiosyncratic component might capture industry-specific layoffs or regional labor market disruptions; in inflation forecasting, it could reflect commodity price shocks or weather-related supply disturbances; and for GDP growth, it may include country-specific fiscal measures or localized political events. Further discussions and examples of these components are provided in Fan et al. (2023). Our approach aligns with the FARM introduced by Fan et al. (2020), Fan et al. (2023), and Fan et al. (2024), as we decompose the individual forecasts into the common factor (the simple average combined forecast) and idiosyncratic components.

In this context, substituting Equation (4) into Equation (1) gives the combined forecast

$$\begin{aligned} f_{c,t+h} &= \sum_{i=1}^{N} \beta_i \left(\bar{f}_{t+h} + d_{i,t+h} \right) \\ &= \bar{f}_{t+h} + \sum_{i=1}^{N} \beta_i d_{i,t+h}, \end{aligned} \tag{5}$$

where the second equality follows from $\sum_{i=1}^{N} \beta_i = 1$. The combined forecast $f_{c,t+h}$ encompasses the simple average combined forecast $\bar{f}_{t+h}$ if it improves forecast performance by selectively incorporating the idiosyncratic components $d_{i,t+h}$.

By re-expressing Equation (5) in terms of forecast errors, we have

$$\bar{u}_{t+h} = \sum_{i=1}^{N} \beta_i d_{i,t+h} + u_{c,t+h}, \tag{6}$$

where $\bar{u}_{t+h} = y_{t+h} - \bar{f}_{t+h}$ and $u_{c,t+h} = y_{t+h} - f_{c,t+h}$. Hence, the weights β_i can be estimated by regressing the simple average forecast errors $\bar{u}_{t+h}$ on the idiosyncratic components $d_{i,t+h}$. Unlike previous studies that estimate combining weights directly using individual forecasts (as in Equation (1)), our method uses information from selected idiosyncratic components to adjust the weights β_i away from equal weights. We treat the simple average $\bar{f}_{t+h}$ as the sole common factor and augment it with selected idiosyncratic components by estimating β_i based on Equation (6).

The motivation for deviating from equal weights is that the simple average combined forecast is optimal only under the restrictive assumption that no idiosyncratic components $d_{i,t+h}$ enter into Equation (6). Our approach allows for heterogeneity in forecast error variances across individual forecasters. By incorporating additional information beyond the simple average, the combining weights are adaptively learned and updated based on these extra signals.

To estimate the combining weights β_i in Equation (6), we employ four penalized regression methods appropriate for high-dimensional settings, where the number of forecasts may be large relative to the sample size. First, we consider the Lasso to select a subset of informative idiosyncratic components $d_{i,t+h}$. This approach operates under the sparsity assumption—that only a small subset of these components contains predictive information for the target variables.

Second, since the Lasso can be inconsistent for variable selection, we employ the Adaptive Lasso (ALasso) of Zou (2006), which can achieve the oracle property. Following Belloni and Chernozhukov (2013), we adopt a post-selection estimation approach for both the Lasso and ALasso to estimate the combining weights β_i using only the selected components.

Third, we employ the component-wise L_2-Boosting procedure of Bühlmann (2006) to estimate β_i in Equation (6). The boosting process begins with the simple average as an initial learner. In each iteration, an idiosyncratic component $d_{i,t+h}$ is selected in the direction that most reduces the forecast error loss, ensuring that each updated combination encompasses the previous one. This process is repeated until a predetermined stopping criterion is reached. Following Hastie et al. (2009), two hyperparameters—the step size (learning rate) and the maximum number of boosting iterations—are set to 0.001 and 3000, respectively, for the empirical applications in Section 4.

Fourth, we consider the Ridge regression of Hoerl and Kennard (1970). Lasso is known to struggle with "weak" signals (i.e., predictors that have negligible influence on the outcome variable). Recently, Shen and Xiu (2025) theoretically demonstrate that the Ridge regression is superior in the presence of many weak signals. The main limitation of Lasso is not its difficulty in distinguishing true signals from noise, but rather its ineffectiveness in penalizing irrelevant signals when many weak signals are present. Since many of the idiosyncratic components $d_{i,t+h}$ may be weak signals—meaning many individual forecasts are only weakly correlated with the forecast target after accounting for the simple average—we also employ the Ridge regression to estimate the combining weights β_i.

Remark 1. *The framework in Section 3 can be extended to cases where individual forecasts share additional common factors beyond the simple average. By decomposing each forecast into the simple average, additional common components, and idiosyncratic components, we explicitly account for the high correlation among individual forecasts. While these additional common factors can be estimated using Principal Component Analysis (PCA), we also consider Supervised PCA (SPCA) following Giglio et al. (2025), given that these factors may be "weak" after accounting for the simple average. This structure addresses the limitations of Diebold and Shin (2019), where reliance on sparsity can be problematic in the presence of additional common factors that drive high correlation among the individual forecasts. For a detailed procedure and empirical results, please refer to the Online Supplementary Appendix.*

4 Applications to Macroeconomic Forecasting

In this section, we conduct real-time out-of-sample forecasting exercises using U.S. macroeconomic data from the Federal Reserve Economic Data (FRED). This large-scale database is ideal for empirical analysis in data-rich environments. Our objective is to demonstrate that the factor-adjusted regularized model (FARM), which treats the simple average as the common factor, yields predictive gains over the simple average combined forecast. We compare the h-month-ahead out-of-sample forecasting performance of our combined forecasts relative to that of the simple-average combined forecast.

Our analysis focuses on forecasting five key U.S. macroeconomic series: real personal income (RPI), the Consumer Price Index for all items (CPIAUCSL), the Personal Consumption Expenditure price index (PCEPI), industrial production (INDPRO), and the civilian unemployment rate (UNRATE). We utilize a monthly dataset from the FRED (hereafter FRED-MD), which includes $N = 119$ predictors and spans the period from January 1960 to December 2019, thus avoiding the disruption of the COVID-19 pandemic. This results in a total of $T = 720$ monthly observations.

We construct $N = 119$ individual forecasts using a simple one-predictor-at-a-time setup. While various complex forecasting methods such as linear regression, regularized regression, factor models, and machine learning techniques could be employed, our aim here is not to maximize individual forecast accuracy. Instead, we aim to demonstrate that the proposed combination method improves upon the simple average by accounting for heterogeneity in individual forecasts.

Specifically, the individual forecast $f_{i,t+h}$ is generated based on the i-th predictor $x_{i,t}$ using the following linear regression model: $y_{t+h} = a_i + b_i x_{i,t} + \epsilon_{i,t+h} = f_{i,t+h} + \epsilon_{i,t+h}$, where $f_{i,t+h} \equiv a_i + b_i x_{i,t}$ is the individual forecast i. The parameters a_i and b_i are estimated by regressing the target y_{t+h} on the i-th predictor $x_{i,t}$ using a rolling window with a sample size of T_1. Following the spirit of Huang et al. (2022), $f_{i,t+h}$ can be viewed as a "scaled" predictor, where the scaling (determined by b_i) captures the predictor's predictive ability.

To construct the combined forecast $f_{c,t+h}$, we estimate the combining weights β_i by applying the FARM framework introduced in Section 3. For a comprehensive comparison against existing approaches, we also consider a simple linear first-order autoregressive model, AR(1), and the GR approach. For GR, the target variable y_{t+h} is regressed on the full set of individual forecasts in Equation (1).

The analysis further incorporates the DS approach, which involves regressing the simple average forecast errors $\bar{u}_{t+h}$ on the individual forecasts $f_{i,t+h}$ to estimate the coefficients δ_i, as shown in Equation (3). We also include the "partially egalitarian" method (denoted as p-DS), following Diebold and Shin (2019). This method utilizes the Lasso for initial forecast selection (Equation (1)) before estimating Equation (3). To estimate the coefficients δ_i in Equation (3), we employ the Lasso, ALasso, L_2-boosting, and the Ridge regression. While the DS approach employs all available forecasts, the p-DS approach uses only the subset selected in the initial step.

Finally, we consider Gradient Boosting as a non-linear model (denoted as NL) in Equation (1). We employ Gradient Boosted Regression Trees, which sequentially fit a series of regression trees (acting as base estimators) with a maximum of five splits to the residuals of the preceding ensemble, using a learning rate of 0.001.

The main focus is to compare the performance of each method relative to the simple average. For the out-of-sample comparison, the full sample (T observations) is divided into three sub-periods: a training period (T_1), an estimation period (T_2), and an out-of-sample period (T_3). Thus, the total number of observations is $T = T_1 + T_2 + T_3$. The training period is used to construct the individual forecasts $f_{i,t+h}$, while the estimation period is used to estimate the combining weights β_i. We employ a rolling-window procedure over the out-of-sample period (T_3 observations) to compare the forecast performance, which helps mitigate the effect of randomness, noise, or shocks at any single forecast evaluation point.

We set $T_1 = 120$, $T_2 = 480$, and $T_3 = 120$. Note that two observations are lost due to data transformations required to make the series stationary, along with an additional h observations needed for constructing h-month-ahead forecasts in the training period. Following McCracken and Ng (2016), we take the first differences of the logarithms of RPI and INDPRO, the second differences of the logarithms of CPIAUCSL and PCEPI, and the first differences of UNRATE.

We evaluate the performance of each forecast combination method by computing its MSFE relative to the simple average. As reported in Table 1, a relative MSFE below one indicates that the method outperforms the simple average. The results are summarized as follows:

1. FARM addresses the forecast combination puzzle by improving upon the simple average combined forecast through the inclusion of idiosyncratic components $d_{i,t+h}$. In particular, the improvement in the MSFE ratios between the simple average and FARM is statistically significant for CPIAUCSL, PCEPI, and UNRATE at shorter horizons ($h = 1, 2$), according to the test of Diebold and Mariano (1995). These improvements, however, become less significant at the longer forecast horizon ($h = 3$).

2. The inclusion of selected idiosyncratic components via Lasso, ALasso, or L_2-Boosting is preferable to the inclusion of all idiosyncratic components via Ridge. In particular, a comparison between L_2-Boosting and Ridge within the FARM framework reveals that L_2-Boosting yields lower MSFE ratios in most cases.

3. FARM outperforms the AR(1) model in that its MSFE relative to the simple average is lower than that of the AR(1). This implies that incorporating heterogeneity through idiosyncratic components provides greater gains in forecast performance than relying solely on the history of the target variable.

4. Given that the simple average is treated as the sole common factor, the DS and FARM approaches share a similar underlying structure. However, FARM remains highly competitive; while the p-DS approach improves upon the DS approach through initial forecast selection, FARM yields even better performance across a wide range of cases by effectively incorporating informative idiosyncratic components.

5. Furthermore, with the exception of CPIAUCSL, PCEPI, and UNRATE at $h = 3$, the relative MSFE of FARM remains lower than that of NL. This suggests that FARM is at least as competitive as the non-linear Gradient Boosting method.

6. Finally, as expected, the GR method yields relative MSFEs significantly greater than one across all five series. This result reflects the inefficiency of applying ordinary least squares without regularization in high-dimensional settings, where the number of forecasts to combine is large.

5 Conclusion

This paper addresses the long-standing forecast combination puzzle—the empirical observation that complex weighting schemes often struggle to outperform the simple average combined forecast. We apply the factor-adjusted regularized framework for forecast combinations by treating the simple average as the primary common factor and leveraging the information from idiosyncratic components.

Our empirical findings demonstrate that the proposed method outperforms the simple average by selectively incorporating idiosyncratic components, thereby helping to resolve the forecast combination puzzle. This result underscores the importance of accounting for idiosyncratic components. Our method can be initialized with any combined forecast and iterated until no further improvement is achieved.

References

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

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Beyhum, J. and Striaukas, J. (2024). Testing for sparse idiosyncratic components in factor-augmented regression models. *Journal of Econometrics*, 244(1):105845.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics*, 34:559–583.

Chan, F. and Pauwels, L. L. (2018). Some theoretical results on forecast combinations. *International Journal of Forecasting*, 34(1):64–74.

Chan, Y. L., Stock, J. H., and Watson, M. W. (1999). A dynamic factor model framework for forecast combination. *Spanish Economic Review*, 1(2):91–121.

Claeskens, G., Magnus, J. R., Vasnev, A. L., and Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3):754–762.

Clemen, R. T. and Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Diebold, F. X. and Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, 6(4):503–508.

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

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72(4):1107–1125.

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–

157.

Elliott, G. and Liao, J. (2025). Combining forecasts-on why averaging beats optimal linear weights. University of California, San Diego. https://econweb.ucsd.edu/~gelliott/Combination.pdf.

Fan, J., Ke, Y., and Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, 216(1):71–85.

Fan, J., Lou, Z., and Yu, M. (2024). Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, 119(546):1076–1088.

Fan, J., Masini, R. P., and Medeiros, M. C. (2023). Bridging factor and sparse models. *The Annals of Statistics*, 51(4):1692–1717.

Figlewski, S. and Urich, T. (1983). Optimal aggregation of money supply forecasts: Accuracy, profitability and market efficiency. *The Journal of Finance*, 38(3):695–710.

Frazier, D. T., Covey, R., Martin, G. M., and Poskitt, D. (2023). Solving the forecast combination puzzle. *arXiv:2308.05263*.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1):108–121.

Giglio, S., Xiu, D., and Zhang, D. (2025). Test assets and weak factors. *The Journal of Finance*, 80(1):259–319.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Green, K. C. and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8):1678–1685.

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

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: data mining, inference, and prediction*. Springer.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Huang, D., Jiang, F., Li, K., Tong, G., and Zhou, G. (2022). Scaled PCA: A new approach to dimension reduction. *Management Science*, 68(3):1678–1695.

Liu, L., Hao, X., and Wang, Y. (2024). Solving the forecast combination puzzle using double shrinkages. *Oxford Bulletin of Economics and Statistics*, 86(3):714–741.

Matsypura, D., Thompson, R., and Vasnev, A. L. (2018). Optimal selection of expert forecasts with integer programming. *Omega*, 78:165–175.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Poncela, P., Rodríguez, J., Sánchez-Mangas, R., and Senra, E. (2011). Forecast combination through dimension reduction techniques. *International Journal of Forecasting*, 27(2):224–237.

Rossi, B. (2021). Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature*, 59(4):1135–1190.

Rossi, B. and Sekhposyan, T. (2013). Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, 177(2):199–212.

Shen, Z. and Xiu, D. (2025). Can machines learn weak signals? No. w33421. National Bureau of Economic Research.

Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355.

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

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

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Table 1: MSFEs Relative to the Simple Average Combined Forecast

		RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE	RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE	RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE
				$h=1$					$h=2$					$h=3$		
AR(1)		1.103 (0.882)	1.131 (0.997)	1.206 (1.000)	1.220 (0.998)	0.961 (0.234)	1.093 (0.851)	1.147 (0.998)	1.211 (1.000)	1.222 (0.999)	0.962 (0.233)	1.101 (0.881)	1.127 (0.997)	1.204 (1.000)	1.219 (0.999)	0.950 (0.184)
GR		1.631 (0.999)	2.064 (1.000)	1.516 (0.978)	1.539 (1.000)	1.175 (0.891)	1.446 (1.000)	1.936 (1.000)	1.660 (0.995)	1.512 (0.993)	1.140 (0.744)	1.662 (1.000)	1.898 (0.989)	1.855 (1.000)	1.482 (0.996)	1.462 (0.998)
DS	Lasso	1.075 (0.990)	0.945 (0.312)	0.894 (0.160)	0.927 (0.179)	0.884 (0.141)	1.002 (0.524)	1.078 (0.739)	1.061 (0.686)	1.041 (0.653)	0.861 **(0.056)**	1.015 (0.723)	1.022 (0.595)	1.134 (0.820)	0.989 (0.437)	0.890 **(0.061)**
	ALasso	1.004 (0.719)	0.881 **(0.078)**	0.823 **(0.013)**	1.068 (0.978)	0.916 **(0.087)**	1.011 (0.969)	1.000 (0.503)	0.995 (0.400)	1.075 (0.979)	0.949 (0.169)	1.000 (0.517)	1.032 (0.946)	1.022 (0.841)	1.069 (0.996)	1.150 (0.997)
	L_2-Boosting	0.976 **(0.010)**	0.994 (0.143)	0.986 **(0.004)**	0.945 (0.127)	0.930 **(0.007)**	0.975 **(0.010)**	1.007 (0.885)	0.989 (0.191)	0.947 (0.107)	0.930 **(0.005)**	0.984 **(0.095)**	0.990 **(0.089)**	0.990 (0.117)	0.944 (0.122)	0.941 **(0.010)**
	Ridge	1.006 (0.777)	0.920 **(0.007)**	0.902 **(0.009)**	1.018 (0.947)	0.884 **(0.029)**	1.007 (0.841)	0.937 **(0.022)**	0.918 **(0.054)**	1.028 (0.930)	0.875 **(0.004)**	0.995 (0.269)	0.972 **(0.050)**	0.990 (0.366)	1.029 (0.999)	0.975 **(0.015)**
p-DS	Lasso	0.990 (0.301)	0.952 (0.339)	0.879 (0.128)	0.974 (0.345)	0.887 (0.150)	0.987 **(0.083)**	1.115 (0.812)	1.040 (0.636)	1.044 (0.681)	0.851 **(0.041)**	0.998 (0.384)	1.011 (0.550)	1.124 (0.808)	0.974 (0.308)	0.915 **(0.094)**
	ALasso	0.983 (0.179)	0.914 (0.144)	0.856 **(0.017)**	0.999 (0.493)	0.905 **(0.080)**	0.988 **(0.086)**	0.986 (0.349)	1.028 (0.863)	1.051 (0.920)	0.949 (0.190)	0.997 (0.365)	1.1017 (0.646)	1.034 (0.704)	1.016 (0.704)	1.085 (0.922)
	L_2-Boosting	0.976 **(0.010)**	0.991 **(0.070)**	0.987 **(0.004)**	0.945 (0.126)	0.930 **(0.007)**	0.975 **(0.010)**	1.007 (0.885)	0.990 (0.194)	0.947 (0.107)	0.930 **(0.005)**	0.984 **(0.095)**	0.990 **(0.089)**	0.990 (0.117)	0.944 (0.122)	0.940 **(0.009)**
	Ridge	0.982 (0.158)	0.867 **(0.038)**	0.864 **(0.034)**	0.991 (0.387)	0.860 **(0.039)**	0.995 (0.256)	0.954 (0.261)	0.983 (0.422)	0.982 (0.295)	0.828 **(0.005)**	0.996 (0.278)	0.967 (0.244)	1.021 (0.600)	1.031 (0.974)	0.932 **(0.046)**
FARM	Lasso	1.034 (0.689)	0.990 (0.464)	0.827 **(0.038)**	0.998 (0.488)	0.931 (0.193)	1.052 (0.886)	1.005 (0.518)	1.117 (0.775)	1.154 (0.927)	0.914 (0.165)	1.071 (0.956)	1.133 (0.894)	1.020 (0.650)	1.035 (0.695)	0.961 (0.217)
	ALasso	0.966 (0.327)	0.885 **(0.077)**	0.902 **(0.024)**	0.935 (0.151)	0.862 **(0.054)**	0.986 (0.123)	0.962 (0.157)	1.024 (0.683)	0.985 (0.356)	0.859 **(0.027)**	1.015 (0.806)	1.085 (0.827)	1.018 (0.633)	0.985 (0.298)	1.005 (0.539)
	L_2-Boosting	0.961 (0.241)	0.883 **(0.004)**	0.866 **(0.002)**	0.985 (0.340)	0.877 **(0.019)**	1.014 (0.697)	0.929 **(0.070)**	0.932 **(0.031)**	1.086 (0.972)	0.877 **(0.020)**	1.000 (0.504)	1.005 (0.557)	1.014 (0.631)	0.961 (0.114)	0.948 **(0.086)**
	Ridge	1.007 (0.829)	0.932 **(0.053)**	0.913 **(0.007)**	1.016 (0.946)	0.926 **(0.082)**	0.998 (0.395)	0.997 **(0.043)**	0.952 **(0.026)**	1.005 (0.577)	0.927 **(0.050)**	1.006 (0.797)	1.001 (0.926)	1.003 (0.997)	1.013 (0.913)	0.969 **(0.033)**
NL	Gradient Boosting	1.110 (0.975)	0.988 **(0.034)**	0.994 (0.171)	0.993 (0.387)	0.938 **(0.039)**	1.092 (0.940)	1.012 (0.907)	0.993 (0.182)	1.011 (0.664)	0.934 **(0.034)**	1.103 (0.969)	0.986 **(0.037)**	0.992 (0.149)	0.998 (0.468)	0.927 **(0.022)**

[1] Table 1 reports the h-month-ahead out-of-sample forecasting performance for five U.S. macroeconomic series relative to the simple average combined forecast, where $h = 1, 2, 3$. A relative MSFE below one indicates that the combined forecast from a given method outperforms the simple average; such values are highlighted in blue. The p-value reported in parentheses is based on the Diebold and Mariano (1995) test. Values in bold indicate significance at the 10% level.

[2] The "DS" panel reports the results of the egalitarian method of Diebold and Shin (2019) using Lasso, ALasso, L_2-Boosting, and Ridge regression, while the "p-DS" panel reports the results of the partially egalitarian method of Diebold and Shin (2019) using the same set of regularization techniques.

Online Supplementary Appendix to "Improving the Simple Average Combined Forecast via Factor-Adjusted Regularization"

Saerom Lee* and Tae-Hwy Lee†

March 22, 2026

The main paper treats the simple average as the sole common factor. In this Online Supplementary Appendix, we extend our framework to incorporate additional common factors beyond the simple average. This decomposition allows each individual forecast $f_{i,t+h}$ to be expressed as:

$$f_{i,t+h} = \bar{f}_{t+h} + \boldsymbol{\lambda}_i' \boldsymbol{g}_{t+h} + d_{i,t+h}, \tag{A.1}$$

where $\boldsymbol{g}_{t+h}$ is an r-dimensional vector of additional common factors, and $\boldsymbol{\lambda}_i$ is the corresponding r-dimensional vector of loadings. Each forecast $f_{i,t+h}$ thus consists of $(r+1)$ common components, including the simple average $\bar{f}_{t+h}$.

In this context, the combined forecast is obtained by substituting Equation (A.1) into the linear combination:

$$\begin{aligned} f_{c,t+h} &= \sum_{i=1}^{N} \beta_i \left(\bar{f}_{t+h} + \boldsymbol{\lambda}_i' \boldsymbol{g}_{t+h} + d_{i,t+h} \right) \\ &= \bar{f}_{t+h} + \boldsymbol{\alpha}' \boldsymbol{g}_{t+h} + \sum_{i=1}^{N} \beta_i d_{i,t+h}, \end{aligned} \tag{A.2}$$

*Department of Economics, The University of Toledo, Toledo, OH 43606. E-mail: saeromlee3390@gmail.com.
†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu.

where $\boldsymbol{\alpha} \equiv \sum_{i=1}^{N} \beta_i \boldsymbol{\lambda}_i'$ is an r-dimensional vector representing the effect of the additional common factors $\boldsymbol{g}_{t+h}$ on the combined forecast. The contribution of the additional common factors $\boldsymbol{g}_{t+h}$ to the target variable y_{t+h} can be consistently quantified by regressing the deviation $(y_{t+h} - \bar{f}_{t+h})$ on $\boldsymbol{g}_{t+h}$. This is because the additional common factors $\boldsymbol{g}_{t+h}$ and the idiosyncratic components $d_{i,t+h}$ are uncorrelated.

We apply Principal Component Analysis (PCA) to estimate the additional common factors ($\boldsymbol{g}_{t+h}$), the corresponding loadings ($\boldsymbol{\lambda}_i$), and the number of factors (r) following Stock and Watson (2002), Bai (2003), and Bai and Ng (2002). However, a key challenge is that these additional common factors may be "weak" after accounting for the simple average. To address this, we also consider Supervised PCA (SPCA) of Giglio et al. (2025) as an alternative to standard PCA. Unlike the Ridge regression, which we use to handle potentially weak idiosyncratic components ($d_{i,t+h}$) in the main paper, SPCA is employed here to address potentially weak common components ($\boldsymbol{g}_{t+h}$) by selecting a subset of forecasts to strengthen the extracted factors.

The SPCA procedure is iterative, combining supervised selection and factor extraction. It begins by computing the univariate correlation between each $d_{i,t+h}$ (where $d_{i,t+h} = f_{i,t+h} - \bar{f}_{t+h}$ from the previous subsection) and the target variable y_{t+h}. Components $d_{i,t+h}$ with sufficiently high absolute correlation are selected. The first estimated factor is then constructed using PCA applied to these selected components. These two steps are repeated r times. In each subsequent iteration, the selection and extraction are performed on the residuals of y_{t+h} and $d_{i,t+h}$, both obtained after removing the portion explained by the common factors estimated in the previous iteration. The number of additional common factors, r, is determined using 3-fold cross-validation to minimize the MSFE. After r iterations, the residuals of $d_{i,t+h}$ are purely idiosyncratic components.

We rearrange Equation (A.2) and rewrite it in terms of forecast errors as follows:

$$\tilde{u}_{t+h} = \sum_{i=1}^{N} \beta_i d_{i,t+h} + u_{c,t+h}, \tag{A.3}$$

where $\tilde{u}_{t+h} \equiv y_{t+h} - \left(\bar{f}_{t+h} + \boldsymbol{\alpha}'\boldsymbol{g}_{t+h}\right)$ represents the forecast error after removing all common components, including the simple average, and $u_{c,t+h}$ is the combined forecast error. We estimate β_i by regressing the forecast errors $\tilde{u}_{t+h}$ on the idiosyncratic components $d_{i,t+h}$.

This structure addresses the limitation of the Diebold and Shin (2019). While Diebold and Shin (2019) also use a forecast encompassing approach, their reliance on sparsity in the coefficients δ_i can be problematic when individual forecasts contain additional common factors beyond the simple average, leading to high correlation among the individual forecasts $f_{i,t+h}$. In contrast, our method explicitly accounts for these additional common factors, thereby allowing for sparsity in the true weights β_i because the resulting idiosyncratic components ($d_{i,t+h}$) are weakly correlated.

To evaluate the performance of this extended framework, we apply it to the empirical applications presented in Section 4 of the main paper. Table A1 reports the results for the case involving common factors beyond the simple average. FARM improves the simple average combined forecast by incorporating selected idiosyncratic components, as shown in the results presented in Table 1 of the main paper. FARM2 and FARM3 further enhance the simple average forecast by selectively incorporating both additional common factors and idiosyncratic components. In particular, FARM2 uses PCA to estimate the additional common factors, while FARM3 employs SPCA. We set the maximum number of additional common components, r, to 2 for these empirical applications.

Interestingly, incorporating additional common components does not further improve the simple average forecast. This finding suggests that FARM1 ($r = 0$) may be sufficient, implying that FARM2 or FARM3 may be unnecessary in these particular applications. One explanation is that the simple average already captures the most important common factor shared across individual forecasts, while estimating additional common factors and their corresponding coefficients ($\boldsymbol{\alpha}$) may introduce estimation errors that offset any potential gains. However, a comparison between FARM2 and FARM3 shows that FARM3 often outperforms FARM2. This result highlights that any additional common factors beyond the simple average in these applications are likely weak.

Table A1: MSFEs Relative to the Simple Average Combined Forecast

		RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE	RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE	RPI	CPIAUCSL	PCEPI	INDPRO	UNRATE
		$h = 1$					$h = 2$					$h = 3$				
FARM	Lasso	1.034	0.990	0.827	0.998	0.931	1.052	1.005	1.117	1.154	0.914	1.071	1.133	1.020	1.035	0.961
		(0.689)	(0.464)	**(0.038)**	(0.488)	(0.193)	(0.886)	(0.518)	(0.775)	(0.927)	(0.165)	(0.956)	(0.894)	(0.650)	(0.695)	(0.217)
	ALasso	0.966	0.885	0.902	0.935	0.862	0.986	0.962	1.024	0.985	0.859	1.015	1.085	1.018	0.985	1.005
		(0.327)	**(0.077)**	**(0.024)**	(0.151)	**(0.054)**	(0.123)	(0.157)	(0.683)	(0.356)	**(0.027)**	(0.806)	(0.827)	(0.633)	(0.298)	(0.539)
	L_2-Boosting	0.961	0.883	0.866	0.985	0.877	1.014	0.929	0.932	1.086	0.877	1.000	1.005	1.014	0.961	0.948
		(0.241)	**(0.004)**	**(0.002)**	(0.340)	**(0.019)**	(0.697)	**(0.070)**	**(0.031)**	(0.972)	**(0.021)**	(0.504)	(0.557)	(0.631)	(0.114)	**(0.086)**
	Ridge	1.007	0.932	0.913	1.016	0.926	0.998	0.997	0.952	1.005	0.927	1.006	1.001	1.003	1.013	0.969
		(0.829)	**(0.053)**	**(0.007)**	(0.946)	**(0.082)**	(0.395)	**(0.043)**	**(0.027)**	(0.577)	**(0.050)**	(0.797)	(0.926)	(0.997)	(0.913)	**(0.033)**
FARM2	Lasso	0.968	0.983	0.900	1.214	1.123	1.024	1.153	1.167	1.186	1.051	1.048	1.154	1.020	1.132	1.162
		(0.375)	(0.434)	**(0.071)**	(0.993)	(0.991)	(0.859)	(0.869)	(0.815)	(0.964)	(0.853)	(0.885)	(0.897)	(0.671)	(0.964)	(1.000)
	ALasso	0.965	0.892	0.931	1.117	1.053	0.997	0.960	0.981	1.159	1.069	1.014	1.070	1.006	1.037	1.099
		(0.264)	**(0.050)**	**(0.039)**	(0.963)	(0.950)	(0.348)	(0.194)	(0.321)	(0.951)	(0.982)	(0.743)	(0.855)	(0.552)	(0.737)	(1.000)
	L_2-Boosting	0.954	0.902	0.911	1.078	1.081	1.010	0.970	0.969	1.144	1.055	0.999	1.020	1.027	1.019	1.137
		(0.168)	**(0.002)**	**(0.002)**	(0.985)	(1.000)	(0.748)	(0.207)	(0.135)	(0.988)	(0.979)	(0.481)	(0.799)	(0.786)	(0.671)	(1.000)
	Ridge	1.005	0.952	0.993	1.063	1.102	1.005	1.001	1.002	1.070	1.095	1.003	1.003	1.003	1.065	1.105
		(0.727)	**(0.021)**	(0.167)	(1.000)	(1.000)	(0.734)	(0.677)	(0.707)	(0.967)	(1.000)	(0.717)	(0.752)	(0.743)	(0.995)	(1.000)
FARM3	Lasso	1.008	1.119	0.916	1.028	1.046	1.010	1.070	1.103	1.160	0.938	1.036	1.132	1.010	1.010	1.000
		(0.546)	(0.868)	**(0.076)**	(0.699)	(0.824)	(0.662)	(0.778)	(0.815)	(0.948)	(0.162)	(0.872)	(0.868)	(0.583)	(0.567)	(0.497)
	ALasso	0.971	0.969	0.867	0.971	1.024	1.024	0.951	1.037	0.983	0.925	1.029	1.074	1.015	0.996	1.010
		(0.350)	(0.213)	**(0.004)**	(0.262)	(0.765)	(0.879)	(0.135)	(0.808)	(0.327)	**(0.074)**	(0.893)	(0.798)	(0.623)	(0.438)	(0.623)
	L_2-Boosting	0.957	0.935	0.928	0.998	1.009	1.017	0.949	0.963	1.090	0.934	0.998	1.003	1.010	0.973	0.977
		(0.207)	**(0.023)**	**(0.007)**	(0.462)	(0.647)	(0.807)	**(0.071)**	**(0.052)**	(0.989)	**0.046**	(0.406)	(0.542)	(0.600)	(0.149)	(0.189)
	Ridge	1.006	0.947	0.938	1.034	1.016	1.000	0.999	0.977	1.023	0.963	1.006	1.001	1.002	1.026	0.988
		(0.814)	**(0.041)**	**(0.008)**	(0.998)	(0.775)	(0.519)	(0.171)	**(0.078)**	(0.857)	**(0.087)**	(0.815)	(0.909)	(0.967)	(0.991)	(0.171)

[1] Table A1 reports the h-month-ahead out-of-sample forecasting performance for five U.S. macroeconomic series relative to the simple average combined forecast, where $h = 1, 2, 3$. A relative MSFE below one indicates that the combined forecast from a given method outperforms the simple average; such values are highlighted in blue. The p-value reported in parentheses is based on the Diebold and Mariano (1995) test. Values in bold indicate significance at the 10% level.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.