

Ensembles and combinations: using multiple models to improve forecasts

1 Simple Combination

The idea of combining forecasts is derived from the simple portfolio diversification argument, which is a risk management strategy with an obvious intuition: do not put all eggs into one basket. As early as 1963, [Barnard \(1963\)](#) provided an empirical justification of forecast combination by making a simple average of two forecasts and obtained smaller forecast errors than individual forecasts. [Bates and Granger \(1969\)](#) developed a general analysis and further explored more possibilities for forecast combination by extending a simple average to a weighted combination. It is often considered to be the seminal work on point forecast combination. Ensemble is a machine learning paradigm which, similar to combination, uses multiple models to solve the same problem. It is difficult to trace the beginning of the history of ensemble forecasting. However, it is clear that ensemble techniques have become a hot topic in various fields, especially weather forecasting, since the 1990s. An overview on ensemble forecasting with a focus on weather forecasting is given by [Leutbecher and Palmer \(2008\)](#). A genealogy to depict the scientific roots of ensemble forecasting from several fundamental lines of research is provided in [Lewis \(2005\)](#). Combinations and ensembles have been used successfully in the forecasting community to produce improved forecasts on average by combining multiple models.

Considerable literature has accumulated over the years regarding the way in which individual models are combined. A unanimous conclusion is that simple combination schemes are hard to beat ([Clemen, 1989](#); [Stock and Watson, 2004](#); [Lichtendahl and Winkler, 2020](#)). More specifically, simple combination methods which ignore past information regarding the precision of individual forecasts and correlations between forecast errors work reasonably well relative to more sophisticated combination schemes (see [Clemen, 1989](#)). [Timmermann \(2006\)](#) concisely summarized the reasons for this success by the importance of parameter estimation error—that is, simple combination schemes do not require estimating parameters such as combining weights based on forecast errors, thus avoiding parameter estimation error that often exists in the weighted combination.

The vast majority of studies on combining multiple models has dealt with point forecasting,

even though point forecasts generally provide insufficient information for decision making. The simple average based on equal weights is the most widely used combining method (see [Bunn, 1985](#); [Clemen and Winkler, 1986](#); [Stock and Watson, 2003](#)) and its advantages are summarized in [Palm and Zellner \(1992\)](#). [Clemen \(1989\)](#) provided a review and annotated bibliography containing over two hundred articles regarding the combination of forecasts, and then addressed the issue that the arithmetic mean often dominates more refined forecast combination. This empirical fact has been first stated as the “forecast combination puzzle” in [Stock and Watson \(2004\)](#). [Timmermann \(2006\)](#) pointed out that the good average performance of the simple average depends strongly on model instability and the ratio of forecast error variances associated with different forecasting models. Furthermore, some theoretical explanation of the forecast combination puzzle are offered in later studies (see [Smith and Wallis, 2009](#); [Claeskens et al., 2016](#)). The simple average of individual forecasts is sensitive to some large outliers, so the median and the trimmed mean of the panel of forecasts have been suggested due to its advantage of being less affected by extreme values (e.g., [Chan et al., 1999](#); [Stock and Watson, 2004](#); [Jose and Winkler, 2008](#); [Genre et al., 2013](#)).

Compared to various refined combination approaches and advanced machine learning algorithms, simple combinations appear outdated and uncompetitive in the big data era. However, the results from the recent M4 competition ([Makridakis et al., 2020](#)) show that simple combinations can achieve fairly good forecasting performance and still be competitive. Specifically, a simple equal-weights combination achieved the third best performance for yearly time series ([Shaub, 2019](#)) and a median combination of four models achieved sixth place for the point forecasts ([Petropoulos and Svetunkov, 2020](#)). Therefore, simple combinations provide a natural benchmark to measure the effectiveness of the newly proposed weight estimation algorithms (e.g., [Makridakis and Hibon, 2000](#); [Makridakis et al., 2020](#); [Montero-Manso et al., 2020](#); [Kang et al., 2020](#); [Wang et al., 2021](#)).

There has been a growing interest in using combination methods for probabilistic forecasting. Probabilistic forecasts differ from point forecasts in the sense that there is an associated probability related to the reported values, providing more comprehensive information about uncertainties of the future load than single-valued forecasts.

References

- Barnard, G. A. (1963), ‘New methods of quality control’, *Journal of the Royal Statistical Society. Series A* **126**(2), 255.
- Bates, J. M. and Granger, C. W. J. (1969), ‘The combination of forecasts’, *The Journal of the Operational Research Society* **20**(4), 451–468.
- Bunn, D. W. (1985), ‘Statistical efficiency in the linear combination of forecasts’, *International Journal of Forecasting* **1**(2), 151–163.
- 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. (1989), ‘Combining forecasts: A review and annotated bibliography’, *International Journal of Forecasting* **5**(4), 559–583.
- Clemen, R. T. and Winkler, R. L. (1986), ‘Combining economic forecasts’, *Journal of Business & Economic Statistics* **4**(1), 39–46.
- 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.
- Jose, V. R. R. and Winkler, R. L. (2008), ‘Simple robust averages of forecasts: Some empirical results’, *International Journal of Forecasting* **24**(1), 163–169.
- Kang, Y., Hyndman, R. J. and Li, F. (2020), ‘GRATIS: GeneRAtIng TIme Series with diverse and controllable characteristics’, *Statistical analysis and data mining* **13**(4), 354–376.
- Leutbecher, M. and Palmer, T. N. (2008), ‘Ensemble forecasting’, *Journal of Computational Physics* **227**(7), 3515–3539.
- Lewis, J. M. (2005), ‘Roots of ensemble forecasting’, *Monthly Weather Review* **133**(7), 1865–1885.
- Lichtendahl, K. C. and Winkler, R. L. (2020), ‘Why do some combinations perform better than others?’, *International Journal of Forecasting* **36**(1), 142–149.
- URL:** <https://www.sciencedirect.com/science/article/pii/S0169207019301475>

- Makridakis, S. and Hibon, M. (2000), ‘The M3-Competition: results, conclusions and implications’, *International Journal of Forecasting* **16**(4), 451–476.
- Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2020), ‘The M4 competition: 100,000 time series and 61 forecasting methods’, *International Journal of Forecasting* **36**(1), 54–74.
- Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J. and Talagala, T. S. (2020), ‘FFORMA: Feature-based forecast model averaging’, *International Journal of Forecasting* **36**(1), 86–92.
- Palm, F. C. and Zellner, A. (1992), ‘To combine or not to combine? issues of combining forecasts’, *Journal of Forecasting* **11**(8), 687–701.
- Petropoulos, F. and Svetunkov, I. (2020), ‘A simple combination of univariate models’, *International Journal of Forecasting* **36**(1), 110–115.
- Shaub, D. (2019), ‘Fast and accurate yearly time series forecasting with forecast combinations’, *International Journal of Forecasting* .
- 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. (2003), How did leading indicator forecasts perform during the 2001 recession?
- 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.
- Timmermann, A. (2006), Chapter 4 forecast combinations, in G. Elliott, C. W. J. Granger and A. Timmermann, eds, ‘Handbook of Economic Forecasting’, Vol. 1, Elsevier, pp. 135–196.
- Wang, X., Kang, Y., Petropoulos, F. and Li, F. (2021), ‘The uncertainty estimation of feature-based forecast combinations’, *Journal of the Operational Research Society* .