# Ensembles and combinations: using multiple models to improve forecasts

# 1 Research beginnings

- Goal

- Application

## 1.1 Combinations

Barnard (1963) first proposed an empirical justification of forecast combination by making a simple average of two forecasts in forecasting airline passenger numbers. It was observed that a simple average of two forecasts outperformed each of them. This was the first work to provide an analysis for the problem of point forecast combination.

Bates and Granger (1969) further explored more possibilities for forecast combination by extending the simple average to a weighted combination, which examined various weight determination methods yield from past errors (MSE). It is concluded that combining forecasts with some independent information improves forecast accuracy.

## 1.2 Ensembles

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. Lewis (2005) provided a genealogy to depict the scientific roots of ensemble forecasting from several fundamental lines of research.

Ensemble forecasting, which aims at quantifying the forecast uncertainty and generating probability forecasts, achieved its practical implementation in weather forecasting as a response to the limitations of deterministic forecasting without taking account of the initial condition perturbations. Leutbecher and Palmer (2008) discussed the sources of uncertainty in weather forecasting: uncertainties in the observations, used to initialise the predictions, and in the forecast models themselves.

## 2  Weather a training dataset is required

### 2.1  Combinations

Barnard (1963) provided the arithmetic mean of the two individual forecasts in the application of forecasting airline passenger data (not required).

Bates and Granger (1969) extended the simple average to a weighted combination which aims to give higher weight to better forecaster. Several ways of determining combining weights based on past errors are introduced for combining point forecasts (not required).

Newbold and Granger (1974) examined various possibilities of combining three univariate time series forecasting methods (Box-Jenkins, Holt-Winters and stepwise autoregression) over a large collection of economic time series. Their results suggested that a combined forecast generally performs well if the weights calculated based on the relative precision of individual forecasts ignore the effects of correlations among errors (not required). Winkler and Makridakis (1983) further confirmed their conclusions (not required).

Granger and Ramanathan (1984) considered a linear combination of forecasts by performing a regression model having the actual value as the response variable and the individual forecasts as the explanatory variables. They determined combining weights through a constrained ordinary least squares (OLS) and demonstrated its superiority in fitting and out-of-sample forecasting (not required).

Diebold and Pauly (1990) developed a weight estimation technique by using Bayesian shrinkage techniques to incorporate prior information into unrestricted regression-based forecast combination (not required).

Hibon and Evgeniou (2005) proposed a criterion to chose individual methods and combinations using the smallest average sMAPE over validation periods and successfully applied this criterion on the data of the M3-competition (not required).

Collopy and Armstrong (1992) provided one of the pioneering studies in combining forecasts using characteristics. They proposed a rule-based forecasting procedure to develop 99 rules for forecast combination based on 18 features obtained judgementally (not required).

Vokurka et al. (1996) proposed a rule-based expert forecasting system to automatically identify

time series features and build a weighted combination of three individual methods (not required).

Prudêncio and Ludermir (2004) used machine learning techniques to define the best linear combination of methods. Specifically, they use the Multi-Layer Perceptron (MLP) network as the learner to associate the time series features and the combining weights for two individual methods (required for building MLP learner).

Lemke and Gabrys (2010) created a feature pool describing both the time series and the pool of individual forecasting methods, and then applied different meta-learning approaches to improve forecasting performance using a ranking-based combination (required for building different meta-learning models).

Andrawis et al. (2011) carefully selected the individual forecasting models and combined them by simple average (not required).

Montero-Manso et al. (2020) employed 42 time series features to estimate the optimal combination weights of nine forecasting methods based on extreme gradient boosting (required for training XGBoost using features and errors).

Kang et al. (2020) proposed GeneRAting TIme Series with diverse and controllable characteristics (GRATIS) with the use of mixture autoregressive (MAR) models. GRATIS can be used as a tool for simulating a diverse set of time series data which serves as a training dataset for forecast combination (required).

Li et al. (2020) transformed time series into images and then extracted features from these images for forecast combination (required for training an XGBoost model to produce weights for individual methods).

## 2.2 Ensembles

Cordeiro and Neves (2009) developed the Boost.EXPOS approach which forecasts time series by combining the use of exponential smoothing methods with the bootstrap resampling technique for time series data (not required).

Bergmeir et al. (2016) presented a bagging exponential smoothing method which ensembles exponential smoothing models estimated on the bootstrapped series (not required).

Petropoulos et al. (2018) implemented six forecasting strategies, four of which focus on isolating each one of the three sources of uncertainty (data, parameter, and model uncertainty) to separately explore the benefits of bagging for time series forecasting for each one of them (not required).

Taieb and Hyndman (2014) proposed a forecasting strategy which adjusts autoregressive forecasts with a direct strategy using a boosting procedure at each horizon (not required).

Barrow and Crone (2016) transformed a time series into a dataset. The dataset consists of several pairs of observations comprised of a lagged autoregressive (AR) realisation vector and a further realisation. A generic AdaBoost algorithm is applied in the dataset to obtain combined forecasts (not required).

Ribeiro and dos Santos Coelho (2020) compared several models to explore the predictive capability of regression ensembles. They considered stacked generalization by training a meta-model which determines combining weights by using individual forecasts as an input set (a training dataset is created by the approach of leave-one-out cross-validation for training meta-model).

# 3 Data resampling for modeling

# 4 Forecast method pool

- Homogeneous methods

- Heterogeneous methods

- Selection of individual forecasting models

# 5 Forecasting structure

- Parallel forecasting

- Sequential forecasting

# 6 Main types of ensembles

- Bagging

- Boosting

- Stacking

# 7 Simple aggregation methods

- Such as mean, trimmed mean, median, trimmed median, ...

# 8 Weight determination

- Data base

- Information used

- Weather to consider correlations in estimating combining weights

- Fixed weights VS time-varying weights

# 9 Uncertainty tackling

- Three sources of uncertainty: model, parameter, and data uncertainty

# 10 Contribution to forecast improvement

- How much do combinations and ensembles contribute to the forecast improvement?

# 11 Point forecasting

# 12 Probabilistic forecasting

- Residual simulation

- Combining interval forecasts

- Combining quantile forecasts

- Combining probability forecasts


# 13    Forecast calibration

- Perfectly calibrated, poorly calibrated (overconfident, underconfident)

- Overfit and overconfident forecasts


# References

Andrawis, R. R., Atiya, A. F. and El-Shishiny, H. (2011), 'Forecast combinations of computational intelligence and linear models for the nn5 time series forecasting competition', *International Journal of Forecasting* **27**(3), 672–688.

Barnard, G. A. (1963), 'New methods of quality control', *Journal of the Royal Statistical Society. Series A (General)* **126**(2), 255–258.

Barrow, D. K. and Crone, S. F. (2016), 'A comparison of AdaBoost algorithms for time series forecast combination', *International Journal of Forecasting* **32**(4), 1103–1119.

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

Bergmeir, C., Hyndman, R. J. and Benítez, J. M. (2016), 'Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation', *International Journal of Forecasting* **32**(2), 303–312.

Collopy, F. and Armstrong, J. S. (1992), 'Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations', *Management Science* **38**(10), 1394–1414.

Cordeiro, C. and Neves, M. (2009), 'Forecasting time series with BOOT. EXPOS procedure', *REVSTAT-Statistical Journal* **7**(2), 135–149.

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

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

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

Kang, Y., Hyndman, R. J. and Li, F. (2020), 'GRATIS: GeneRAting TIme Series with diverse and controllable characteristics', *Statistical Analysis and Data Mining: The ASA Data Science Journal* **13**(4), 354–376.

Lemke, C. and Gabrys, B. (2010), 'Meta-learning for time series forecasting and forecast combination', *Neurocomputing* **73**(10-12), 2006–2016.

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.

Li, X., Kang, Y. and Li, F. (2020), 'Forecasting with time series imaging', *Expert Systems with Applications* **160**, 113680.

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.

Newbold, P. and Granger, C. W. (1974), 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of the Royal Statistical Society: Series A (General)* **137**(2), 131–146.

Petropoulos, F., Hyndman, R. J. and Bergmeir, C. (2018), 'Exploring the sources of uncertainty: Why does bagging for time series forecasting work?', *European Journal of Operational Research* **268**(2), 545–554.

Prudêncio, R. and Ludermir, T. (2004), Using machine learning techniques to combine forecasting methods, *in* 'Australasian Joint Conference on Artificial Intelligence', Springer, pp. 1122–1127.

Ribeiro, M. H. D. M. and dos Santos Coelho, L. (2020), 'Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series', *Applied Soft Computing* **86**, 105837.

Taieb, S. B. and Hyndman, R. (2014), Boosting multi-step autoregressive forecasts, *in* 'International Conference on Machine Learning', PMLR, pp. 109–117.

Vokurka, R. J., Flores, B. E. and Pearce, S. L. (1996), 'Automatic feature identification and graphical support in rule-based forecasting: a comparison', *International Journal of Forecasting* **12**(4), 495–512.

Winkler, R. L. and Makridakis, S. (1983), 'The combination of forecasts', *Journal of the Royal Statistical Society: Series A (General)* **146**(2), 150–157.