MUNI ECON

Artificial Intelligence in Finance

Forecast combinations

Štefan Lyócsa

Department of Finance, Faculty of Economics and Administration

December 11, 2024

Outline for Section 1

The problem

Is there a single '*best model*'?

Not sure ..., but even if, are we able to find it? That might not be easy Petropoulos et al. (2018, [8]; Kourentzes et al (2019,[5]):

- **Data** uncertainty - we observe a specific finite sample.
- **Parameter** uncertainty - with finite sample we are unsure about the estimated parameters.
- **Model** uncertainty - do we have the right model (specification)?

The solution

In 1906, sir F. Galton observed that an average of 787 estimates of ox's weight is quite close to the ox's actual weights (Surowiecki, 2005 [10]) $\rightarrow$ wisdom in crowds?

Solution to the '*one ring that rule them all*' problem is to combine multiple forecasts (from different forecasting models) into a prediction $\rightarrow$ **combination forecasts** or **ensemble forecast**. Why it might work?

- **Partial/incomplete overlap** of information from different models.
- **Structural breaks** in model parameters.
- **Parameter uncertainty** reduction, e.g. forecast bias and variance reduction.

The solution

The research has since evolved with studies of Bates and Granger (1969, [2]) or Timmermann, (2006,[11]) being highly influential. The solution leads to the following tasks:

- **How to combine** model forecasts?
  - Point forecasts:
    - Simple models.
    - Linear combinations - optimal weights, regression based, performance based.
    - Non-linear combinations.
  - Probabilistic forecasts.
- **What models** to **combine**?
  - The 'good' ones.
  - Diverse models.

Outline for Section 2

Simple ensembles

Let $F_{m,i}$ be a forecast from model $m = 1, 2, ..., M$ of observation $i = 1, 2, ..., N$. Combination that ignores past performance and dependence between forecasts is a **simple combination forecast**.

- **Average** forecast:

$$F_i = M^{-1} \sum_{m=1}^{M} F_{m,i} \tag{1}$$

- **Median** forecast (Ox), let $F_{(r),i}$ denote ordered forecasts from the lowest ($r = 1$) to the highest ($r = M$) value:

$$F_i = \begin{cases} 2^{-1} \left(F_{\left(\frac{M}{2}\right),i} + F_{\left(\frac{M+1}{2}\right),i} \right) & 2M + 1 : M \in \mathbb{Z} \\ F_{\frac{M}{2},i} & 2M : M \in \mathbb{Z} \end{cases} \tag{2}$$

- **Winsorized** average:

$$F_i = (M - 2\lfloor \alpha M \lfloor)^{-1} \sum_{r=\lfloor \alpha M \rfloor}^{M-\lfloor \alpha M \rfloor} F_{(r),i} \tag{3}$$

Simple ensembles

Why such simple methods may work, according to Pam and Zellner (1992, [7]):

- Averaging reduces bias and variance of errors.
- No need to estimate weights $\rightarrow$ **what method to employ**?
- No need to worry about **weight estimation error**.
- No need for extra observations for calibrating the weights.

Makridakis et al (2020, [6]) reports that these methods still offer **competitive forecasts**. However, a known **downside** is that the result depends on the 'quality' of the models employed. Why?

Linear combinations: Optimal weights

Instead of using equally weighted averages, one wants to find weights w_m that minimize the variance of the combined forecasts $\rightarrow$ '*optimal*' weights (Bates and Granger, 1969, [2]).

If the individual forecasts are **unbiased** and forecast **error variances** are **stable**, the combined forecast is **unbiased**.

Optimal weights are given via a restricted OLS estimation of (Granger and Ramanathan, 1984, [3]):

$$F_i = \sum_{m=1}^{M} w_m F_{m,i} + \epsilon_i, \text{ s.t.} \sum_{m=1}^{M} = 1 \quad (4)$$

Linear combinations: regression based forecasts

An unrestricted model is possible:

$$F_i = \sum_{m=1}^{M} w_m F_{m,i} + \epsilon_i \tag{5}$$

Often, forecasts are highly correlated $\rightarrow$ collinearity issues $\rightarrow$ RIDGE.

Some of the forecasts might work well for a specific **subspace** of the **features space**. You can **add features** to the model:

$$F_i = \sum_{m=1}^{M} w_m F_{m,i} + \sum_{k=1}^{K} \beta_k X_{i,k} + \epsilon_i, \tag{6}$$

Linear combinations: principal components

Often, forecasts are highly correlated $\rightarrow$ collinearity issues $\rightarrow$ application of the principal components regression - **PRC** (see Stock and Watson, 2004, [9]). In the **first step** extracting first S principal components, $P_s, s = 1, 2, ..., S \leq M$ could:

- Describe the forecasts well (high explained variance), while reducing the noise.
- Principal components are orthogonal.

The later is useful in the **second step**, where the following model is being estimated via OLS:

$$F_i = \sum_{s=1}^{S} w_s P_{s,i} + \epsilon_i, \tag{7}$$

The true out-of-sample forecast is found by using estimated loadings to estimate the principal components ($P_{s,i}$) and using weights w_s to arrive at the final forecast.

Performance-based combinations

Shouldn't forecasts from more accurate models receive higher weights?

Regression-based forecasts might enforce this principle. A popular alternative given by Bates and Granger (1969, [2]) is a **performance-based weight**. Let V_i be the predicted value:

$$w_m = \frac{\bar{l}_m^{-1}}{\sum_{k=1}^{M} \bar{l}_k^{-1}} \quad (8)$$

with $\bar{l}_m$ being the sum (average) of forecast loss values, i.e. mean square error $N^{-1}\sum_i^N (V_i - F_m)^2$, mean absolute error $N^{-1}\sum_i^N |(V_i - F_m)|$.

Performance-based combinations

Averages (sum) are sensitive to outliers. A more **robust** and **less sensitive** to outliers approach is based on the **ranking** of individual forecasts (Wang et al, 2023 [12]), that was suggested by Aiolfi and Timmermann (2006, [1]).
Let R_m denote the rank of the ordered average loss values ($\bar{l}_m$), where $R_m = 1$ corresponds to the model with lowest errors and $R_m = M$ to the model with highest errors. The **rank performance-based weights** are given:

$$w_m = \frac{R_m^{-1}}{\sum_{k=1}^{M} R_k^{-1}} \quad (9)$$

Non-linear combination forecasts

Many regression based models can be enhanced by non-linear counterparts, where we map individual forecasts to observed values using non-linear methods:

- Bagged tree.
- Random forest.
- Boosted tree.
- Neural networks.

Outline for Section 3

Identify good models: K-means

'Many could be better than all' (Zhou et al 2002, [13])

In case of many forecasts, one can use **k-means** to find groups of forecasts that tend to produce similar forecasting accuracy; for a given loss function (e.g. mean square error, mean absolute error,...).

Aiolfi and Timmermann (2006, [1]) suggest to create such clusters and use simple combination forecasts:

- Average.
- Median.
- Trimmed average.

or performance-based forecasts.

Identify good models: Model confidence set

'Many could be better than all' (Zhou et al 2002, [13])

With not that many models (otherwise it takes long on larger samples) one could rely on statistical test.

Using the **model confidence set** approach of Hansen et al (2011, [4]) identify a set of superior models. One could use simple combination forecasts with the set of superior models:

- Average.
- Median.
- Trimmed average.

or performance-based forecasts.

Identify good models: Diverse models

'Many could be better than all' (Zhou et al 2002, [13])

It is usually useful to combine individual forecasts from models that are similarly accurate, but forecast are **not correlated** $\rightarrow$ **diverse models**.

This should motivate you to try many different types of models:

- different model types,
- different information (feature) sets.

Even if you do not have a clear winner, combinations might prove to be beneficial with diverse models.

[1] Marco Aiolfi and Allan Timmermann. “Persistence in forecasting performance and conditional combination strategies”. In: *Journal of Econometrics* 135.1-2 (2006), pp. 31–53.

[2] John M Bates and Clive WJ Granger. “The combination of forecasts”. In: *Journal of the operational research society* 20.4 (1969), pp. 451–468.

[3] Clive WJ Granger and Ramu Ramanathan. “Improved methods of combining forecasts”. In: *Journal of forecasting* 3.2 (1984), pp. 197–204.

[4] Peter R Hansen, Asger Lunde, and James M Nason. “The model confidence set”. In: *Econometrica* 79.2 (2011), pp. 453–497.

[5] Nikolaos Kourentzes, Devon Barrow, and Fotios Petropoulos. “Another look at forecast selection and combination: Evidence from forecast pooling”. In: *International Journal of Production Economics* 209 (2019), pp. 226–235.

[6] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. "The M4 Competition: 100,000 time series and 61 forecasting methods". In: *International Journal of Forecasting* 36.1 (2020), pp. 54–74.

[7] Franz C Palm and Arnold Zellner. "To combine or not to combine? Issues of combining forecasts". In: *Journal of Forecasting* 11.8 (1992), pp. 687–701.

[8] Fotios Petropoulos, Rob J Hyndman, and Christoph Bergmeir. "Exploring the sources of uncertainty: Why does bagging for time series forecasting work?" In: *European Journal of Operational Research* 268.2 (2018), pp. 545–554.

[9] James H Stock and Mark W Watson. "Combination forecasts of output growth in a seven-country data set". In: *Journal of forecasting* 23.6 (2004), pp. 405–430.

[10] James Surowiecki. *The wisdom of crowds*. Anchor, 2005.

[11] A Timmerman. *Forecast combination, Handbook of Economic Forecasting, Edited by: Graham Elliott, Clive WJ Granger and Allan Timmermann*. 2006.

[12] Xiaoqian Wang et al. “Forecast combinations: an over 50-year review”. In: *International Journal of Forecasting* 39.4 (2023), pp. 1518–1547.

[13] Zhi-Hua Zhou, Jianxin Wu, and Wei Tang. “Ensembling neural networks: many could be better than all”. In: *Artificial intelligence* 137.1-2 (2002), pp. 239–263.

MUNI
ECON

Artificial Intelligence in Finance

Forecast combinations

Štefan Lyócsa

Department of Finance, Faculty of Economics and Administration

December 11, 2024

MASARYK UNIVERSITY