Revisiting the forecast combination puzzle with different data types: An empirical study

A research proposal submitted for the degree of

Bachelor of Commerce (Honours)

by

Xiefei Li

30204232

xlii0145@student.monash.edu

Supervisor: David T. Frazier

David.frazier@monash.edu

Department of Econometrics and Business Statistics

Monash University

Australia

April 2023

Contents

Introduction

1.1 Research Question and Objective

This paper aims to demonstrate the presence of the forecast combination puzzle in various settings besides the time series domain and to examine the general solution to the forecast combination puzzle empirically. The combination puzzle refers to the empirical finding that the simple average combination method often out-performs sophisticated combination methods. Over the past 50 years, the empirical study undertaken so far has been limited, in that most attention have been focused on different time series datasets. Therefore, it is necessary to explore whether the forecast combination puzzle is present in other data types. Furthermore, the general solution for the forecast combination puzzle is still lack of empirical support. This study will be then extended to examine the application of the general explanation and solution proposed by Zischke et al. (2022) and Frazier et al. (2023) through a forecasting accuracy test.

1.2 Motivation

The forecasting accuracy is always a concern when forecasts are used in the decision-making. Under the classical frequentist approach, forecasters often choose only one "best model" to mimic the actual data generating process of the interested variable and then use it to predict future values. However, that single model could be misleading as it may not capture all important features of the data. The idea of combining multiple estimates of unknown interests already exists before the well-known seminal work conducted by Bates and Granger (1969). They popularized the use of forecast combination for optimal forecasts with a number of combination techniques. The dramatic improvements in forecast accuracy through flexible combination methods attract increasing attention and contributions among researchers from different fields, both theoretical and empirical [Clemen (1989);T06]; see Wang et al. (2022) for a modern literature review over the past 50 years.

In short, forecast combinations involve producing point or density forecasts, and then combining them based on a rule or a weighting scheme. This process can incorporate more independent and unique characteristics of the true data generating process to mitigate different sources of uncertainties. However, issues could arise with arbitrary or careless implementation. One surprising phenomenon in many empirical study, coined by Stock and Watson (2004), is the so-called “forecast combination puzzle” - “theoretically sophisticated weighting schemes should provide more benefits than the simple average from forecast combination, while empirically the simple average has been continuously found to dominate more complicated approaches to combining forecasts” (Wang et al., 2022) (see the section 2.6 for more details and examples). This counter-intuitive result is widely discovered in the time series settings, what will happen when working with datasets such as surveys of professional forecasters, dynamic panels, and pure cross-sectional?

Lastly, a general solution of explaining the forecast combination puzzle is still under development and lack of public acceptance. If the forecast combination puzzle occurs in every setting, it is then essential to explore its cause and to support the theory with empirical evidence. Frazier et al. (2023) demonstrated that, in theory, the cause of the puzzle is the way researchers produce the forecast, named a “two-step” approach in the paper. The constituent model forecasts are determined at first with estimated parameters and the unknown weights are then estimated conditional on all the estimates in the first stage. Due to the unawareness and the dimensionality of combining all unknown parameters, this two-step approach is commonly studied and used in the literature, e.g. HM07; Geweke and Amisano (2011); Gneiting and Ranjan (2013); BS16. Frazier et al. (2023) further claims that the forecast combination puzzle can be avoided when unknown parameters and weights are estimated in one step, namely a “one-step” approach, when feasible. In other words, if forecasts are produced by estimating parameters and weights simultaneously, the sophisticated weighting schemes should (asymptotically) be superior. This new finding relies on the investigation of forecast combination performance conducted by Zischke et al. (2022) in terms of the one- and two-step approach. In this paper, I will use some real data to empirically support all their ideas along with a measure of forecast accuracy test by examine the null hypothesis of *no inferior forecast performance*.

Methodology

The first goal is to construct linear density forecast combinations with any two parametric models for a given data. On top of point forecasts, using density forecast can benefit forecasters or decision markers with a broader view of understanding the target variable and potential risks (see the section 2.6.1. of Petropoulos et al. (2022) for related contributions). The weighting scheme is to maximize the log predictive score function, which is comprised of two selected forecast densities. The procedure refers to the two-step approach mentioned before. The results should not be surprising that the forecast combinations indeed improve the forecast accuracy via assessing the log predictive score function.

Next goal is to estimate unknown parameters of constituent models and the weight in a single step. This one-step approach is expected to have a better performance than the two-step approach stated above by conducting the forecast accuracy test.

Before explaining further details, the following notations will be used throughout the paper. The dataset with T observations will be divided proportionally into two parts, an in-sample period R and an out-of-sample period P. The realization of a target variable y at time t is denoted as y_t. Its future value after the in-sample period is denoted as y_{R+h} where h is the forecast horizon. $\mathcal{F}_t$, the information set at time t, is consists of all observed (and known) realizations of y up to time t, i.e., $\mathcal{F}_t = \{y_1, y_2, .., y_t\}$.

2.1 Forecast Combination Method

For the first step, I will estimate unknown parameter of each constituent model by the Maximum Likelihood Estimation. Estimates will then be held fixed and substituted into their corresponding probability density function.

With the idea of linear pooling (Bates and Granger, 1969; Hall and Mitchell, 2007; Geweke and Amisano, 2011), the linear combinations of two predictive densities $f^{(t)}$ will be constructed with two constituent predictive densities $f_1^{(t)}$ and $f_2^{(t)}$:

$$f^{(t)}(y) = wf_1^{(t)}(y) + (1-w)f_2^{(t)}(y) \tag{2.1}$$

where $f_1^{(t)}(y)$ and $f_2^{(t)}(y)$ are assumed to follow the normal distributions but with different means and variances, h is the future value after the in-sample period (R), and w is the weight allocated to the first model. Through this construction, the sum of two weights is implied to be 1, which is necessary and sufficient for the combination to be a density function(Geweke and Amisano, 2011).

More specifically, $f_1^{(t)}(y) = f_1(y_t|\mathcal{F}_{t-1}) = N\{y_t; \mu_1, \sigma_1^2\}$ and $f_2^{(t)}(y) = f_2(y_t|\mathcal{F}_{t-1}) = N\{y_t; \mu_2, \sigma_2^2\}$. $N\{x; \mu, \sigma^2\}$ denotes the normal probability density function evaluated at value x with mean μ and variance σ^2. Given $\mathcal{F}_{t-1}$, the conditional mean and conditional variance should be used.

2.2 Evaluation of Weighted Forecast Combinations

This refers to the second step where I can estimate the weight given the aforementioned estimates for parameters.

Following section 1.1 and 1.2 of Geweke and Amisano (2011), w is estimated by maximizing the log predictive score function:

$$\sum_{h=1}^{P} log\Big[wf_1(y_{R+h}) + (1-w)f_2(y_{R+h})\Big] \tag{2.2}$$

where $f(y_{R+h}) = f(y_{R+h}|\mathcal{F}_{R+h-1})$.

2.3 A Motivating Example

2.3.1 Data

Reconsidering the example in section 3 of Geweke and Amisano (2011), I use the daily Standard and Poor's (S&P) 500 index from February 11, 2013 to February 10, 2023 (10 years in total) retrieved via the FRED (2023). Total 2519 (T) available observations are partitioned into two periods with rough proportion. The in-sample period contains the first 60% of the data ($R = 1511$), which is used for estimating unknown parameters in each model. The rest 40% ($P = 1008$) becomes the out-of-sample period for further evaluation.

2.3.2 Model Specification

For a simple illustration purpose, I choose five basic models to study the performance of two-model pools:

1. Model 1: An ARIMA(1,1,1) model with an intercept for the natural logarithm of S&P 500.
2. Model 2: An ETS(M,N,N) model for the S&P 500.
3. Model 3: An ETS(M,A,N) model for the S&P 500.

All error terms are assumed to be independent and normally distributed with mean zero and variance σ_m^2 for $m = 1, 2, 3$.

4. Model 4: A linear regression model for the S&P 500 with a trend regressor and errors, follow an ARIMA(1,0,0) process.
5. Model 5: A linear regression model for the natural logarithm of S&P 500 with a trend regressor and errors follow an ARIMA(1,0,0) process.

Both error terms in the ARIMA model are assumed to be independent and normally distributed with mean zero and variance σ_m^2 for $m = 4, 5$.

All unknown parameters in each model are estimated by maximizing the likelihood function with the in-sample period data. Estimated values are held fixed for the density evaluations. For each model, I generate the predictive densities at every future time point of S&P 500 ($h = 1, 2, ..., P$) given that all information before that point is known.

Preliminary Results

3.1 Tables

3.2 Figures

Appendix

All analyses were performed using R Statistical Software (R version 4.2.1 (2022-06-23))

Packages used are `tidyverse` (Wickham et al., 2019), `dplyr` (Wickham et al., 2023), and `fpp3` (Hyndman, 2023).

$$
\begin{aligned}
M_1 : log(y_t) &= \phi_{0,1} + log(y_{t-1}) + \phi_{1,1} log(y)_{t-1} + \theta_{1,1}\epsilon_{t-1} + \epsilon_{t,1} \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_1^2) \\
M_2 : y_t &= \ell_{t-1,2}(1 + \epsilon_{t,2}) \\
\ell_{t,2} &= \ell_{t-1,2}(1 + \alpha_2 \epsilon_{t,2}) \\
M_3 : y_t &= (\ell_{t-1} + b_{t-1})(1 + \epsilon_{t,3}) \\
\ell_t &= \ell_{t-1}(1 + \alpha \epsilon_{t,2}) \\
M_4 : y_t &= \\
M_5 : y_t &=
\end{aligned}
$$

$$
y_t - y_{t-4} = \beta(x_t - x_{t-4}) + \gamma(z_t - z_{t-4}) + \phi_1(y_{t-1} - y_{t-5}) + \Theta_1 \varepsilon_{t-4} + \varepsilon_t \tag{A.1}
$$

Hyndman and Athanasopoulos (2021)

Reference

Bates, JM and CW Granger (1969). The combination of forecasts. *Journal of the operational research society* **20**(4), 451–468. DOI: https://doi.org/10.1057/jors.1969.103.

Clemen, RT (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting* **5**(4), 559–583.

Frazier, DT, R Zischke, GM Martin, and DS Poskitt (2023). Solving the Forecast Combination Puzzle. [In preparation].

FRED (2023). *S&P500*. https://fred.stlouisfed.org/series/SP500#0 (visited on 02/12/2023).

Geweke, J and G Amisano (2011). Optimal prediction pools. *Journal of Econometrics* **164**(1), 130–141. DOI: https://doi.org/10.1016/j.jeconom.2011.02.017.

Gneiting, T and R Ranjan (2013). Combining predictive distributions.

Hall, SG and J Mitchell (2007). Combining density forecasts. *International Journal of Forecasting* **23**(1), 1–13.

Hyndman, R and G Athanasopoulos (2021). *Forecasting: principles and practice, 3rd edition*. OTexts.com/fpp3 (visited on 02/12/2023).

Hyndman, R (2023). *fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition)*. R package version 0.5. https://CRAN.R-project.org/package=fpp3.

Petropoulos, F, D Apiletti, V Assimakopoulos, MZ Babai, DK Barrow, SB Taieb, C Bergmeir, RJ Bessa, J Bijak, JE Boylan, et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*.

Stock, JH and MW Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting* **23**(6), 405–430. DOI: https://doi.org/10.1002/for.928.

Wang, X, RJ Hyndman, F Li, and Y Kang (2022). Forecast combinations: an over 50-year review. *International Journal of Forecasting*. DOI: https://doi.org/10.48550/arXiv.2205.04216.

Wickham, H, M Averick, J Bryan, W Chang, LD McGowan, R François, G Grolemund, A Hayes, L Henry, J Hester, M Kuhn, TL Pedersen, E Miller, SM Bache, K Müller, J Ooms, D Robinson, DP Seidel, V Spinu, K Takahashi, D Vaughan, C Wilke, K Woo, and H Yutani (2019). Welcome to the tidyverse. *Journal of Open Source Software* **4**(43), 1686. DOI: 10.21105/joss.01686.

Wickham, H, R François, L Henry, K Müller, and D Vaughan (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.1. https://CRAN.R-project.org/package=dplyr.

Zischke, R, GM Martin, DT Frazier, and DS Poskitt (2022). The Impact of Sampling Variability on Estimated Combinations of Distributional Forecasts. *arXiv preprint arXiv:2206.02376*. DOI: https://doi.org/10.48550/arXiv.2206.02376.