International Journal of Forecasting 34 (2018) 64–74

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Some theoretical results on forecast combinations

Felix Chan [a,*], Laurent L. Pauwels [b]

[a] School of Economics and Finance, Curtin University, Australia
[b] University of Sydney Business School, Australia

ARTICLE INFO

Keywords:
Forecast combination
Averaging
Optimal weights
Mean squared error

ABSTRACT

This paper proposes a framework for the analysis of the theoretical properties of forecast combination, with the forecast performance being measured in terms of mean squared forecast errors (MSFE). Such a framework is useful for deriving all existing results with ease. In addition, it also provides insights into two forecast combination puzzles. Specifically, it investigates why a simple average of forecasts often outperforms forecasts from single models in terms of MSFEs, and why a more complicated weighting scheme does not always perform better than a simple average. In addition, this paper presents two new findings that are particularly relevant in practice. First, the MSFE of a forecast combination decreases as the number of models increases. Second, the conventional approach to the selection of optimal models, based on a simple comparison of MSFEs without further statistical testing, leads to a biased selection.
© 2017 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Several studies in the forecasting literature have provided insightful discussions of forecast combination (see Timmermann, 2006, for a survey of the literature). However, the theoretical results on forecast combination have been based on many different techniques and assumptions. As a consequence, it is not clear from that body of literature whether there are deeper connections that underlie the theoretical results of these studies. This paper makes three contributions to the forecast combination literature.

First, this paper proposes a framework for the analysis of the theoretical properties of forecast combination. The framework simplifies the derivations of existing results, showing that they can be derived easily. More importantly, the framework provides a solid foundation from which to study forecast combination, and also yields further insights into the theoretical properties of forecast combination. Specifically, this paper provides the necessary and sufficient conditions regarding the optimal weight that minimises the mean squared forecast errors (MSFE). This result also provides the necessary and sufficient conditions under which the simple average is the optimal weight.

Second, this paper establishes the necessary and sufficient conditions required for the forecasting of a single model to outperform the simple average of forecasts, with the forecasting performance being measured in terms of MSFEs. When the difference in MSFEs is defined as a simple bilinear form, it is trivial to demonstrate that these conditions rarely hold in practice, and thus, the result provides a simple theoretical justification for the superiority of forecasts of averaging models over those of single models. The results also suggest that the MSFE of any affine forecast combination decreases as the number of models increases under a mild set of assumptions.

These theoretical results generalise not only the findings of the seminal paper by Bates and Granger (1969), but also the more recent findings of Capistran and Timmerann (2009) and Hsiao and Wan (2014). Hsiao and Wan (2014) impose a multi-factor structure on forecast errors and develop several eigenvector approaches to the combination of forecasts. These authors provide the conditions under which their approaches yield results identical to the regression approach for determining the optimal weight vector,

* Correspondence to: School of Economics and Finance, Curtin University, GPO BOX U1987, Perth, 6845, Australia.
E-mail address: Felix.Chan@cbs.curtin.edu.au (F. Chan).

http://dx.doi.org/10.1016/j.ijforecast.2017.08.005
0169-2070/© 2017 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

as suggested by Granger and Ramanathan (1984). Hsiao and Wan (2014) also provide a necessary and sufficient condition where the simple average is an optimal combination, but this requires the estimation of a scaling constant in cases where the models produce biased forecasts. This paper contributes to the literature by showing that many of the results of Hsiao and Wan (2014) can be derived without needing the multi-factor structure. Thus, we show that their results hold more generally.

Third, this paper presents a theoretical investigation of the forecast combination puzzle, namely, why forecast combination using simple averages often outperforms complicated weighting schemes in terms of MSFEs. This paper provides new insight into this puzzle; in particular, it shows that the MSFE is merely a variance estimator of the forecast errors, and may not be consistent. Thus, selecting the optimal model based on a naïve comparison of MSFE values without further statistical testing will lead to biased results.

This puzzle was first noted by Clemen (1989), and was formally named the "forecast combination puzzle" by Stock and Watson (2004). Recent studies, such as those of Claeskens, Magnus, Vasnev, and Wang (2016) and Smith and Wallis (2009), argue that the forecast combination puzzle is a result of the presence of finite sample errors in the estimation of optimal weights. Specifically, Smith and Wallis (2009) suggest that part of the reason why the simple average forecast combination is superior to the complex weighting scheme with respect to the MSFE is that the weights must be estimated, which results in finite sample errors. Claeskens et al. (2016) follow in the footsteps of Smith and Wallis (2009) and show that, when the weights need to be estimated, the forecast combination is biased, and the variance of the combination is larger than in a fixed-weights case, such as the simple average. Other, earlier explanations that point to the estimation error as the source of the problem include those of Clemen and Winkler (1986), who investigate parameter instability as the underlying cause of error; Hendry and Clements (2004), who consider discrete shifts in the data generating process and misspecified forecasting models; and Aiolfi and Timmermann (2006). However, Elliott (2011) investigates the hypothesis that the gains from combination are outweighed by the estimation error. Furthermore, Elliott (2011) examines the extent of the theoretical gains with respect to optimal forecast combination and provides the conditions under which averaging and optimal combination are equivalent. These results can be derived easily from the proposed framework. Moreover, this framework helps in the identification of another explanation of the forecast combination puzzle. Specifically, it highlights the fact that MSFE can be viewed as an estimator of the forecast variance that is consistent under standard regularity conditions.

The remainder of this paper is organised as follows. Section 2 revisits forecast combinations in the context of the proposed framework, with the forecasts being evaluated using the MSFE as the criterion. It also proposes a simple bilinear form for evaluating the relative efficiency of two forecast combinations. Section 3 uses the proposed framework to derive several theoretical results, including the conditions under which the simple average is the optimal weight. The relative performance difference between the simple average and the estimated optimal weight is also discussed. Specifically, the impacts of finite sample errors for within-sample estimation and out-of-sample forecast evaluation are investigated in detail. The theoretical results are supported further by several Monte Carlo experiments in Section 4. Concluding remarks are presented in Section 5.

2. Framework and assumptions

This section introduces both the framework and the assumptions used to analyse the theoretical properties of forecast combination. Let f_{it} denote a forecast of the variable y_t from model i, where $i = 0, \ldots, k$, at time t. y_t can be expressed as

$$y_t = f_{it} + \nu_{it}, \qquad i = 0, \ldots, k, \tag{1}$$

where ν_{it} are the forecast errors. Without loss of generality, let f_{0t} be the "best" forecast of variable y_t, based on the forecast criterion $g(f_{it}) \geq 0$, such that for all $\epsilon > 0$,

$$\mathbb{P}[g(\nu_{0t}) < \epsilon] > \mathbb{P}[g(\nu_{it}) < \epsilon] \quad \forall i = 1, \ldots, k,$$

where $\mathbb{P}(A)$ denotes the probability of event A. If we let $u_{it} = \nu_{0t} - \nu_{it}$, rearranging Eq. (1) yields

$$f_{it} = y_t - \nu_{0t} + u_{it}, \qquad i = 1, \ldots, k. \tag{2}$$

This framework decomposes the prediction errors ν_{it} into two parts. The first part, ν_{0t}, represents the prediction error of the best model, while the second part, u_{it}, represents the difference in prediction errors between the best model and model i.

Let $\mathbf{Y} = (y_1, \ldots, y_T)'$, $\mathbf{f}_t = (f_{1t}, \ldots, f_{kt})'$, $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)'$, $\mathbf{F}_0 = (f_{01}, \ldots, f_{0T})$ and $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)'$ with $\mathbf{u}_t = (u_{1t}, \ldots, u_{kt})'$, $\boldsymbol{\nu}_t = (\boldsymbol{\nu}_{1t}, \ldots, \boldsymbol{\nu}_{kt})'$ with $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_T)'$, and $\boldsymbol{\nu}_0 = (\nu_{01}, \ldots, \boldsymbol{\nu}_{0T})'$. Eq. (2) can then be written in matrix form as

$$\mathbf{F} = (\mathbf{Y} - \boldsymbol{\nu}_0) \otimes \mathbf{i}' + \mathbf{u}, \tag{3}$$

where $\mathbf{i}$ denotes a $k \times 1$ vector of ones, and $\otimes$ denotes the Kronecker product. Thus, the forecasts for $t = 1, \ldots, T$, based on a linear combination of forecasts from the k models, are

$$\mathbf{Fa} = \mathbf{Y}\mathbf{i}'\mathbf{a} - \boldsymbol{\nu}_0\mathbf{i}'\mathbf{a} + \mathbf{ua}. \tag{4}$$

If $\mathbf{a}$ is an affine combination, i.e., $\mathbf{i}'\mathbf{a} = 1$, then $\boldsymbol{\nu}_0 + \mathbf{ua}$ is a $T \times 1$ vector containing the forecast errors of the forecast combination. If $\mathbf{a}$ is not an affine combination, then $\mathbf{Fa}$ does not produce unbiased forecasts because $\mathbb{E}(\mathbf{Fa}) = \mathbb{E}(\mathbf{Y})\,\mathbf{i}'\mathbf{a}$, under the standard assumptions that $\mathbb{E}(\nu_{0t}) = \mathbb{E}(u_{it}) = 0$ for all i and t. This is why only affine combinations of forecasts are considered.

This framework is sufficiently flexible to produce both simple and complex forecast combination models. For example, Bates and Granger (1969) present a simple combination model of two competing forecasts: $f_{ct} = af_{1t} + (1-a)f_{2t}$, with a forecast error of $\nu_{it} = y_t - f_{it}$, $i = 1, 2$. This simply implies that $\mathbf{f}_t = (f_{1t}, f_{2t})'$ and $\mathbf{a} = (a, 1-a)'$ in Eq. (4), and typically $0 \leq a \leq 1$.

Unless otherwise stated, this paper makes the following assumptions:

Download English Version:

https://daneshyari.com/en/article/5106315

Download Persian Version:

https://daneshyari.com/article/5106315

Daneshyari.com