International Journal of Forecasting 31 (2015) 223–237

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecast combination with outlier protection

Gang Cheng *, Yuhong Yang [1]

313 Ford Hall, 224 Church St SE, Minneapolis, MN 55455, United States

ARTICLE INFO

Keywords:
AFTER
Forecast combination
Outlier protection
Robustness
Loss function
M3-competition

ABSTRACT

Numerous forecast combination schemes with distinct properties have been proposed. However, to the best of our knowledge, there has been little discussion in the literature of the minimization of forecast outliers when combining forecasts. It would appear to have gone unnoticed that robust combining, which often improves the predictive accuracy (under square or absolute error losses) when innovation errors have a tail that is heavier than a normal distribution, may have a higher frequency of prediction outliers. Given the importance of reducing outlier forecasts, it is desirable to seek new loss functions which can achieve both the usual accuracy and outlier-protection simultaneously. In this paper, we propose a synthetic loss function and apply it to a general adaptive combination scheme for the outlier-protective combination of forecasts. Both the theoretical and numerical results support the advantages of the new method in terms of providing combined forecasts with fewer large forecast errors and comparable overall performances.

1. Introduction

Forecasting is used widely and regularly to help with decision making in many areas of our modern life. Because of the availability of a wide range of sources of information and methods, and distinct backgrounds/preferences of the forecasters, multiple forecasts are available for the target variable of interest in many applications. In order to obtain the most accurate forecasts by taking advantage of the different candidate forecasts, the strategy of forecast combination is often applied.

Since the seminal work on forecast combination by Bates and Granger (1969), thousands of research papers have been published on this topic, with various combining schemes. For example, combining via simple averaging (e.g., Stock & Watson, 1999), combining via variance–covariance estimation of the candidate forecasts (e.g., Bates & Granger, 1969), combining via Bayesian model averaging (e.g., Min & Zellner, 1993), combining via regression on candidate forecasts (e.g., Granger & Ramanathan, 1984), and combining via exponential re-weighting (e.g., Yang, 2004) have all been studied. Reviews and discussions of the research results are provided by Clemen (1989), Lahiri, Peng, and Zhao (2013), Newbold and Harvey (2002) and Timmermann (2006).

Loss functions play important roles in forecast combination in two intertwining directions: they may serve as a key ingredient in combination formulas, and they are used to define performance evaluation criteria. Consider forecast combination via ordinary least squares regression, for example: the combining weights of the forecasts are trained by minimizing the sum of the squared errors (the L_2-loss), while the performances of the combined forecasts are often evaluated under either the same loss function or a different one such as the L_1-loss.

Indeed, the use of a loss function in the first direction is found in many popular combination schemes, such as the regression based combination (e.g., Bates & Granger, 1969; Granger & Ramanathan, 1984) and many adaptive/recursive forecast combination schemes (e.g., Wei &

* Corresponding author. Tel.: +1 612 508 5360.
E-mail addresses: chen2285@umn.edu (G. Cheng), yangx374@umn.edu (Y. Yang).
[1] Tel.: +1 612 626 8337; fax: +1 612 624 8868.

http://dx.doi.org/10.1016/j.ijforecast.2014.06.004

Yang, 2012; Yang, 2000, 2004; Zou & Yang, 2004). For instance, the L_1-AFTER of Wei and Yang (2012) uses the cumulative L_1-loss to summarize the historical performances of the candidate forecasts in order to decide the combining weights for predicting the next observation.

The need to use loss functions in the second direction is obvious. The objective of any combination strategy is to provide forecasts that will serve some predefined/predetermined goals better, where these goals are often characterized in terms of loss or utility functions. While the symmetric quadratic loss is used most often in both theoretical and empirical research works, other loss functions have also been explored for forecast combination (see e.g. Chen & Yang, 2007; Elliott & Timmermann, 2004; Pai & Lin, 2005; Wei & Yang, 2012; Zeng & Swanson, 1998). In particular, it is important to study asymmetric evaluation criteria in fields such as economics and finance (see e.g. Christoffersen & Diebold, 1997; Diebold, 2001; Granger & Newbold, 1986; Granger & Pesaran, 2000; West, Edison, & Cho, 1993; Zellner, 1986). In our context, for example, the linex loss, lin–lin loss and asymmetric squared loss functions are discussed in detail by Elliott and Timmermann (2004) as forecast performance evaluation criteria.

In addition to the loss functions mentioned above, the frequency of large forecast errors (larger than some thresholds in the positive or negative directions) is also important, since decisions made for the future based on substantially over- or under-forecasting may lead to severe undesirable consequences. For instance, a severe forecast error for demand may lead to a company's drastic over- or under-production, negatively affecting its profit. In spite of the obvious importance of having a minimal frequency of large forecast errors, to the best of our knowledge, there has been little discussion in the literature on the use of combining strategies while controlling for the occurrence of large forecast errors directly. It is clear that optimization under the L_2- or L_1-loss or other performance measures can have some effect on the control of the frequency of large forecast errors, but the control is not explicit. Thus, we are interested in understanding how the different loss functions perform in forecast combinations with respect to the occurrence of large forecast errors. One seemingly unnoticed phenomenon is that, although the use of the L_1-loss in forecast combination often improves on the L_2-loss in obtaining more accurate forecast combinations, it may have a higher tendency to produce large forecast errors. Therefore, unfortunately, as will be seen, a robust combining method may actually work against the goal of having fewer outliers in the context of forecast combination.

In this paper, we propose a synthetic loss function (denoted the L_{210}-loss) that is a linear combination of the L_2-loss, the L_1-loss, and a smoothed L_0-loss that naturally and smoothly penalizes the occurrence of large forecast errors more directly. It is used to propose a new combination algorithm based on the general AFTER scheme of Yang (2004). We establish oracle inequalities in terms of the L_{210}-loss that show the optimal converging properties of the new AFTER method. Our numerical results also support the advantages of our outlier-protective approach, in terms of reducing the frequency of large forecast errors in the combined forecasts while maintaining a comparable accuracy under both the L_2- and L_1-losses.

It should be pointed out that outlier forecasts can be defined in different ways, e.g., in relation to either other candidate forecasts or the observed value. In this work, an outlier forecast refers to a forecast that is far away from the realized value (i.e., the forecast error is large in absolute value). Forecasts that are drastically different from the majority in a panel of forecasts may also be defined as outliers. Such outliers may or may not be a concern in terms of forecast accuracy.

The plan of this paper is as follows: Section 2 discusses the motivation and the design of the loss function L_{210}, and provides numeric examples demonstrating its efficiency in terms of outlier protection. In Section 3, the L_{210}-loss-based AFTER methods are proposed and examined theoretically. Simulation results evaluating the performance of our new combination approach are presented in Section 4. Real data from the M3-Competition (see e.g. Makridakis & Hibon, 2000) are used in Section 5, and these results also confirm the advantages of our methods. Section 6 concludes the paper. The proofs of the theoretical results are presented in the Appendix.

2. Outlier protective loss functions

2.1. A deficiency of the robust L_1-loss

The L_1-loss is relatively more resistant to occasional outliers. This well-known and useful feature is exploited by Wei and Yang (2012), for example, for robust forecast combination, which results in more accurate forecasts. However, the robustness comes at a cost: the L_1-loss is often less outlier protective, in the sense that, when it is used to compare different forecasts, it may not have a sufficient dislike for forecasters that have a higher frequency of outliers but a comparable (or slightly better) cumulative L_1-loss, because it places a smaller penalty (compared to the L_2-loss, for example) on large forecast errors (outliers). To clarify this matter, examples will be provided after reviewing a framework for comparing loss functions.

2.1.1. Objective comparison of loss functions

The comparison of loss functions is usually entangled with the evaluation criteria used to define better forecasters, which typically involve loss functions. To avoid the difficulty due to the circular reference, Chen and Yang (2004) proposed a methodology for comparing loss functions objectively.

In a time series setting, suppose we have a variable Y with two competing forecasters, 1 and 2. Specifically, $\hat{Y}_{1,i}$ and $\hat{Y}_{2,i}$ are the forecasts for Y_i made by forecasters 1 and 2 respectively at time $i-1$. Let $e_{1,i} = Y_i - \hat{Y}_{1,i}$ and $e_{2,i} = Y_i - \hat{Y}_{2,i}$ be the forecast errors. Suppose that $e_{1,i}$ and $e_{2,i}$ are *i.i.d.* from certain distributions, and let F_1 and F_2 be the cumulative distribution functions of $|e_{1,i}|$ and $|e_{2,i}|$, respectively.

If $F_1(x) \geq F_2(x)$ for all $x \geq 0$ (i.e., forecaster 2 first-order stochastically dominates forecaster 1), then, in theory,