International Journal of Forecasting 32 (2016) 754–762

Contents lists available at ScienceDirect

International Journal of Forecasting

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

The forecast combination puzzle: A simple theoretical explanation

Gerda Claeskens[a], Jan R. Magnus[b,c], Andrey L. Vasnev[d,*], Wendun Wang[e,c]

[a] KU Leuven, Belgium

[b] Vrije Universiteit Amsterdam, The Netherlands

[c] Tinbergen Institute, The Netherlands

[d] University of Sydney, New South Wales, Australia

[e] Econometric Institute, Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Forecast combination
Optimal weights

ABSTRACT

This paper offers a theoretical explanation for the stylized fact that forecast combinations with estimated optimal weights often perform poorly in applications. The properties of the forecast combination are typically derived under the assumption that the weights are fixed, while in practice they need to be estimated. If the fact that the weights are random rather than fixed is taken into account during the optimality derivation, then the forecast combination will be biased (even when the original forecasts are unbiased), and its variance will be larger than in the fixed-weight case. In particular, there is no guarantee that the 'optimal' forecast combination will be better than the equal-weight case, or even improve on the original forecasts. We provide the underlying theory, some special cases, and a numerical illustration.

1. Introduction

When several forecasts of the same event are available, it is natural to try and find a (linear) combination of these forecasts that is the 'best' in some sense. If we define 'best' in terms of the mean squared error and the variances and covariances of the forecasts are known, then optimal weights can be derived. In practice, though, these (co)variances are not known and need to be estimated. This leads to estimated optimal weights and an estimated optimal forecast combination. Empirical evidence and extensive simulations show that the estimated optimal forecast combination typically does not perform well, and that the arithmetic mean often performs better. This empirical fact has become known as the 'forecast combination puzzle'.

The history of the puzzle is elegantly summarized by Graefe, Armstrong, Jones, and Cuzán (2014, Section 4), and Smith and Wallis (2009) made a rigorous attempt to explain it, using simulations and an empirical example. They showed that the effect of the error on the estimation of the weights can be large, thus providing an empirical explanation of the forecast puzzle. Smith and Wallis (2009) use the words 'finite-sample' error, which suggests that this error may vanish asymptotically. However, it is not so easy to find an asymptotic justification for ignoring the noise generated by estimating the weights. To begin with, it is not clear what 'asymptotic' means here. What goes to infinity? The number of forecasts? If so, then the number of weights also goes to infinity. The number of observations underlying the total (but finite) set of forecasts? That would make more sense, but it would be difficult to analyze.

* Correspondence to: Office 4160, Abercrombie Building (H70), The University of Sydney Business School, Sydney, NSW 2006, Australia.
E-mail address: andrey.vasnev@sydney.edu.au (A.L. Vasnev).

http://dx.doi.org/10.1016/j.ijforecast.2015.12.005
0169-2070/

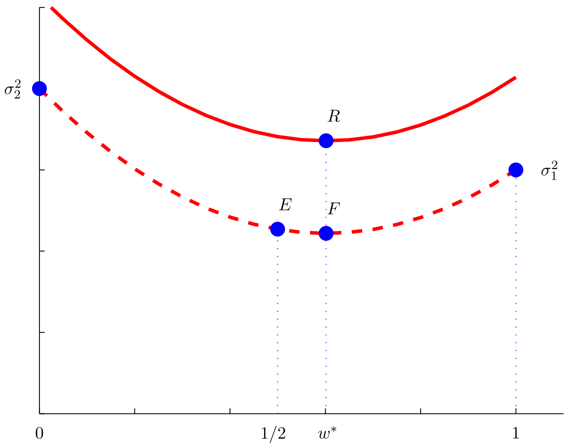

Fig. 1. Variance of forecast combination, in two dimensions: fixed weights (dashed line) and random weights under normality (solid line).

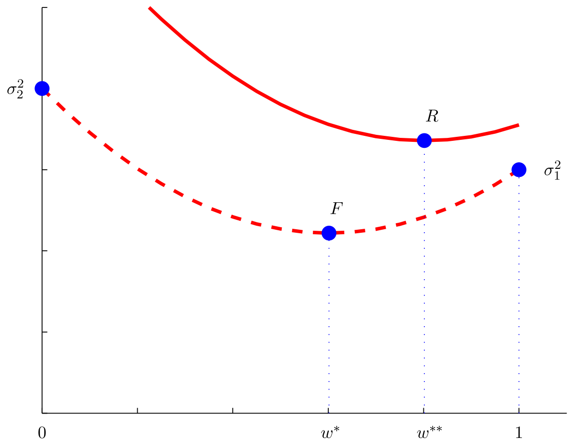

Fig. 2. Variance of forecast combination, in two dimensions: random weights, general case.

In this paper, we provide a theoretical explanation for the empirical and simulation results of Smith and Wallis (2009) and others. The key ingredient of our approach is the specific acknowledgement that the optimal weights should be derived by taking the estimation step into account explicitly. In other words, we view the derivation and estimation of optimal weights as a joint effort, not as two separate efforts. This approach differs from (almost) all previous research, not only the study by Bates and Granger (1969), but also later contributions, important and insightful though they may be, such as those of Elliott (2011), Hansen (2008), Hsiao and Wan (2014), and Liang, Zou, Wan, and Zhang (2011). The separation of the mathematical derivation and statistical estimation can be quite dangerous. However, even though the disadvantages of such separations have been highlighted, they are still quite common in econometrics, and specifically in the model-averaging literature, which explicitly attempts to combine model selection and estimation, so that uncertainty in the model selection procedure is not ignored when reporting properties of the estimates; see for example Magnus and De Luca (2016).

We highlight our main findings by first providing graphical illustrations of the cases of two forecasts, as analyzed by Bates and Granger (1969). Thus, we linearly combine two forecasts of an event μ:

$$y_c = wy_1 + (1 - w)y_2. \tag{1}$$

If the weight w is considered to be fixed, then the forecast combination is unbiased ($\mathrm{E}y_c = \mu$) if the original forecasts are unbiased, and the variance of the combination will be

$$\mathrm{var}(y_c) = w^2\sigma_1^2 + (1 - w)^2\sigma_2^2 + 2w(1 - w)\rho\sigma_1\sigma_2, \tag{2}$$

where σ_1^2 and σ_2^2 are the variances of y_1 and y_2 respectively, and $\rho = \mathrm{corr}(y_1, y_2)$ denotes the correlation.

The variance is a quadratic function of w, as plotted in Fig. 1 (dashed line). At $w = 0$, we obtain σ_2^2; at $w = 1$, we obtain σ_1^2; and at $w = 1/2$, we obtain point E. The optimum F is reached at $w = w^*$, the optimal weight that gives the smallest variance of the forecast combination.

Now suppose that the weights are estimated, so that they are random rather than fixed. In the special case where (y_1, y_2, w) follows a trivariate normal distribution, the combination is biased even when the original forecasts are unbiased, since

$$\mathrm{E}y_c = \mu + \mathrm{cov}(w, y_1 - y_2), \tag{3}$$

and the variance is given by

$$\begin{aligned}\mathrm{var}(y_c) = {} & (\mathrm{E}w)^2\sigma_1^2 + (1 - \mathrm{E}w)^2\sigma_2^2 \\ & + 2(\mathrm{E}w)(1 - \mathrm{E}w)\rho\sigma_1\sigma_2 \\ & + \mathrm{var}(w)\mathrm{var}(y_1 - y_2) + (\mathrm{cov}(w, y_1 - y_2))^2.\end{aligned} \tag{4}$$

In another special case where w is independent of (y_1, y_2), the combination is unbiased and

$$\begin{aligned}\mathrm{var}(y_c) = {} & (\mathrm{E}w)^2\sigma_1^2 + (1 - \mathrm{E}w)^2\sigma_2^2 \\ & + 2(\mathrm{E}w)(1 - \mathrm{E}w)\rho\sigma_1\sigma_2 \\ & + \mathrm{var}(w)\mathrm{var}(y_1 - y_2).\end{aligned} \tag{5}$$

In either case, the variance is shifted upwards, as is shown in Fig. 1 (solid line). The solid line gives the variance as a function of $\mathrm{E}w$, and the optimum is reached at the same point w^* as before, but leading to a higher variance of the forecast combination. We see that, while the equal-weights point at $w = 1/2$ (point E) is not optimal with fixed weights, it has a variance which is smaller than the optimum with estimated weights (point R).

Eqs. (4) and (5) concern special cases (normality and independence, respectively). In general, when the weights are estimated, the combined forecast will be biased, as given in Eq. (3), with its variance given by

$$\begin{aligned}\mathrm{var}(y_c) = {} & (\mathrm{E}w)^2\sigma_1^2 + (1 - \mathrm{E}w)^2\sigma_2^2 \\ & + 2(\mathrm{E}w)(1 - \mathrm{E}w)\rho\sigma_1\sigma_2 \\ & + \mathrm{E}\left[(w - \mathrm{E}w)(y_1 - y_2)\right. \\ & \times \left.((\mathrm{E}w)y_1 + (1 - \mathrm{E}w)y_2 - \mu)\right] \\ & + \mathrm{E}[(w - \mathrm{E}w)^2(y_1 - y_2)^2] \\ & - (\mathrm{cov}(w, y_1 - y_2))^2.\end{aligned} \tag{6}$$

There are now additional terms over and above those in Eqs. (4) and (5), and these shift and distort the fixed-weights curve of Fig. 1, as is illustrated in Fig. 2. The optimal