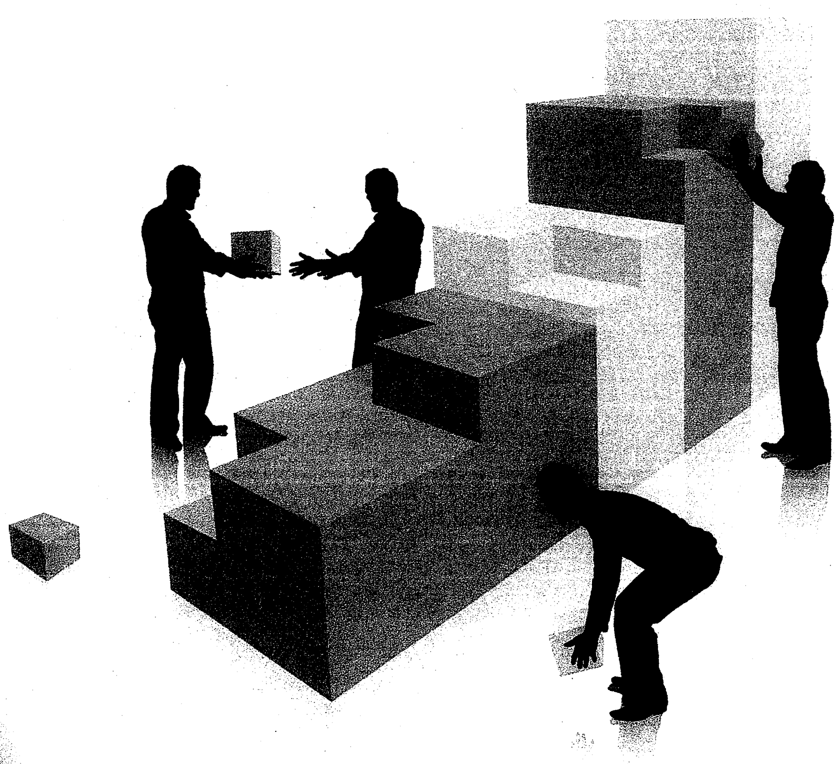
New Evidence on the Value of Combining Forecasts

PAUL GOODWIN

One of the major findings of forecasting research over the last quarter century has been that greater predictive accuracy can often be achieved by combining forecasts from different methods or sources. Combination can be a process as straightforward as taking a simple average of the different forecasts, in which case the constituent forecasts are all weighted equally. Other, more sophisticated techniques are available too, such as trying to estimate the optimal weights that should be attached to the individual forecasts, so that those that are likely to be the most accurate receive a greater weight in the averaging process. Researchers continue to investigate circumstances where combining may well be useful to forecasters and to compare the accuracy of different approaches to combining forecasts.

FORECAST COMBINATION AND THE BANK OF ENGLAND'S SUITE OF STATISTICAL FORECASTING MODELS

George Kapetanios and his colleagues (Kapetanios et al., 2008) have recently evaluated the potential advantages of combining forecasting data at the Bank of England, where quarterly forecasts of inflation and GDP growth are made. The bank has a suite of different statistical forecasting methods available. They include extremely simple approaches, such as the naïve (or random walk) method where the forecasts are equal to the most recent observation. More sophisticated and complex methods in the suite include autoregression, vector-autoregressions (VARs), Markov switching models, factor models, and time-varying coefficient models.



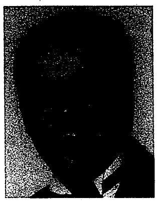
The researchers assessed the value of combining forecasts from the methods available using two different approaches. The first involved taking a simple mean of the forecasts generated by the methods in the suite. The second involved weighting the individual forecasts based upon the Akaike information criterion (AIC). Many commercial forecasting packages report the AIC, which is a measure that takes into account how well a model fits past data but also penalizes the model for complexity, based on the number of parameters it contains. Thus forecasts from relatively simple models that provided a good fit to past observations received a greater weight in the averaging process than more complex or poorer fitting models.

The accuracy of the two types of combined forecasts was assessed over a range of forecast horizons using the relative root mean squared error (RRMSE) statistic. This compares the square root of the sum of squared forecast errors to those of a benchmark forecasting method (in this case, the benchmark was the autoregressive forecast). The researchers reported that "it is striking that forecast performance...is improved when forecasts are combined and the best forecast combinations for both growth and inflation are those based on the [Akaike] information criterion." The Kapetanios group concluded that "combinations of statistical forecasts generate good forecasts of the key macroeconomic variables we are interested in."

Similar benefits of combining have also recently been reported in studies by David Rapach and Jack Strauss (Rapach & Strauss, 2008), who forecast U.S. employment growth, and Jeong-Ryeol Kurz-Kim (Kurz-Kim, 2008), who forecasts U.S. GDP growth. The latter study combined forecasts from the same method (autoregression) that was implemented in different ways.

WHY DID COMBINING WORK?

The researchers suggest a number of reasons. Different models use different sets of information, and each model is likely to represent an incomplete view of the process that is driving the variable of interest. Combined forecasts are therefore able to draw on a wider set of information. In addition, some of the constituent forecasting methods may be biased, in that they consistently forecast too high or too low. When several methods are combined, there is a likelihood that biases in different directions will counteract each other, thereby improving accuracy.



Paul Goodwin is *Foresight*'s Research Column Editor. His earlier columns have covered such topics as tourism forecasting (Summer 2008), new-product forecasting (Spring 2008), supermarket forecasting (Summer 2007), and the value of know-how, training and information sharing (Spring 2007).

TRIMMED MEANS

While the more sophisticated AIC-based weights performed best in the Kapetanios et al. study, the simple mean also did well in both this and the Rapach and Strauss study. The simple mean does have advantages. For one thing, it is easy to implement and explain. It also avoids the need to estimate the optimum set of weights to attach to the forecasts – in many practical circumstances, there may be insufficient data to reliably make these estimates.

However, the simple mean also has the disadvantage of being sensitive to extreme forecasts: if there is an outlying forecast in the set that is being averaged, it will have undue influence on the combined forecast. This has led some researchers (e.g., Armstrong, 2001) to argue that the highest and lowest forecasts should be removed from the set before the mean is calculated. The resulting average is called a *trimmed mean*.

Victor Jose and Robert Winkler (Jose & Winkler, 2008) recently investigated whether trimmed means lead to more accurate combined forecasts. They explored the effects of applying different degrees of trimming (e.g., removing the two highest and two lowest forecasts from the set before averaging, or the three highest and three lowest, and so on). In addition, they evaluated whether an alternative form of averaging, the *Winsorized mean*, was more effective. Rather than removing the highest and lowest forecasts, the Winsorized mean alters their values, making them equal to the highest and lowest forecast values that remain. For example, consider these sales forecasts from five different methods:

When you have access to forecasts from different sources or methods (e.g., different statistical methods or judgmental forecasts from different experts), combining these forecasts is likely to be an effective way of improving accuracy.

23, 34, 47, 53, 86. If we decide to leave off the two "outside" forecasts, our trimmed mean will be the mean of 34, 47, and 53 (i.e., 44.7). In contrast, the Winsorized mean will be the mean of 34, 34, 47, 53, and 53 (i.e., 44.2). It is quickly apparent that these two types of modification only make sense when you have at least three forecasts to work with. Also, the two methods yield differing results only when there are a minimum of five forecasts to combine.

The researchers tested these approaches by combining the forecasts of 22 methods for the 3003 time series from the M3 competition (Makridakis & Hibon, 2000). Additionally, they carried out similar tests on the quarterly nominal GDP forecasts from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters. They found that both trimming and Winsorization yielded slightly more accurate forecasts than the simple mean; they also outperformed all of the individual forecasting methods. There was, however, little to choose between trimming and Winsorization. Moderate degrees of trimming, removing 10 to 30% of the forecasts, seemed to work best. For Winsorization, replacing 15 to 45% of the values appeared to be most effective. I would point out that greater amounts of trimming or replacement yielded greater accuracy when there was more variation in the individual forecasts. This is probably because highly variable sets of forecasts contained extreme values.

CONCLUSIONS

All of this suggests that when you have access to forecasts from different sources or methods (e.g., different statistical methods or judgmental forecasts from different experts), combining these forecasts is likely to be an effective way of improving accuracy. Even using relatively simple combination methods will be enough to yield improvements in many cases. Whatever your area of forecasting, combining forecasts is certainly worth a long, close look.

REFERENCES

Armstrong, J.S. (2001). Combining forecasts. In J.S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers, 417-439.

Jose, V.R. & Winkler, R.L. (2008). Simple robust averages of forecasts: Some empirical results, *International Journal of Forecasting*, 24, 163-169.

Kapetanios, G., Labhard. V. & Price, S. (2008). Forecast combination and the Bank of England's suite of statistical forecasting models, *Economic Modelling*, 24, 772-792.

Kurz-Kim, J-R. (2008). Combining forecasts using optimal combination weight and generalized autoregression, *Journal of Forecasting*, 27, 419-432.

Makridakis, S. & Hibon, M. (2000). The M-3 competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

Rapach, D.E. & Strauss, J.K. (2008). Forecasting US employment growth using forecast combining methods, *Journal of Forecasting*, 27, 75-93.

CONTACT
Paul Goodwin
The Management School, University of Bath
mnspg@management.bath.ac.uk