

Combining Forecasts: The End of the Beginning or the Beginning of the End?*

J. Scott Armstrong

The Wharton School, University of Pennsylvania

Abstract

Research from over 200 studies demonstrates that combining forecasts produces consistent but modest gains in accuracy. However, this research does not define well the conditions under which combining is most effective nor how methods should be combined in each situation. Rule-based forecasting can be used to define these conditions and to specify more effective combinations.

Introduction

“Combining Forecasts” by Clemen (1989) is a milestone on the topic of combining forecasts. His paper summarizes a great amount of research from different fields in a concise and readable manner. It will save time for those who do further research on combined forecasts.

This commentary examines what can be concluded from the research to date, with an emphasis on what has not been learned. I make suggestions about alternative research strategies: ‘Realistic simulation’ and ‘rule-based forecasting’ are especially promising.

Results to Date: The End of the Beginning

As noted by Clemen, past research has produced two primary conclusions, one expected and one surprising. The expected conclusion is that combined forecasts reduce error (in comparison with the average error of the component forecasts). The unexpected conclusion is that the simple average performs as well as more sophisticated statistical approaches.

These conclusions are summarized succinctly by Clemen when he suggests that the forecasting practitioner should “combine and average.” Beyond this conclusion, however, the research has produced few guidelines that are specific enough to be used by forecasters.

Research Strategies: The Beginning of the End

The 209 studies reviewed by Clemen yielded surprisingly little in the way of guidelines for forecasters. Perhaps there has been too much of a reliance on classical statistical procedures. Additional studies of the type already published are unlikely to produce significant gains. In this respect, we are at the beginning of the end.

Alternative strategies are considered here. Among these are (1) meta-analysis, (2) realistic simulations, and (3) rule-based forecasting.

* Support for this paper was provided by the U.S. Navy Personnel Research and Development Center, San Diego, CA 92152, USA.

Meta-analysis

Meta-analysis, the study of studies, involves the statistical analysis of the outcomes of all relevant studies (see Armstrong, 1985, pp. 444-445). Various issues might be addressed by analyzing the results from the 209 studies in Clemen's bibliography.

My recent attempts to use meta-analysis in this situation have led to difficulties. The criteria vary by study and the conditions are not well described. As a result, the coding of the studies has been expensive and reliability has been suspect.

Realistic simulations

An alternative to finding general procedures for combining is to see what works best in each situation. Realistic simulations of the problem faced by the forecaster may help to determine which methods are most effective in the given situation and how they should be combined. Consider the selection of methods to use for a two-year ahead forecast of U.S. GNP. First, a variety of methods would be used to make two-year ahead unconditional forecasts for the U.S. GNP. These would be compared with the actual outcomes. Methods might then be ranked by accuracy, by the extent to which the forecasts differ from one another, and by subjective assessments of the extent to which the methods should differ from one another. The combinations would include the forecasts that differ the most and yet are among the most accurate.

The use of realistic simulations suffers from two defects. First, it does not explain why the various methods are expected to be most useful. Furthermore, if the situation changes, it will take some time until these changes are integrated into the system. Nevertheless, I expect that realistic simulations will improve upon existing procedures.

Rule-based forecasting

Rule-based forecasting (Collopy and Armstrong, 1989) incorporates information from experts and from prior research. The procedure calls for the development of empirically validated and fully disclosed rules for the selection and combination of methods.

The forecaster can provide information about the situation to the rule-based forecasting system. For example, a forecaster might put more weight on an econometric forecast than an extrapolation forecast in a situation where the firm plans to make substantial changes in its marketing procedures.

Much of the research to date has discussed the combination of forecasts from different *models*. It is difficult to interpret the results from a combination of models. Inasmuch as each model contains different elements, it is not always clear why the combinations contribute to increased accuracy.

In our initial development of rule-based forecasting systems, we found that experts prefer to think about the *elements* of the forecasting methods rather than thinking about models. For example, when dealing with extrapolation methods, they think in terms of current level, trend, and seasonal factors. Different combinations can be used for each element of the forecast. Certain measures might be combined to improve the estimate of the starting value, while other approaches might be combined to forecast the trend.

The process of combining the various elements might benefit from intervention by different experts for the different elements. For example, some experts may have substantial knowledge of the current situation; they may know about a recent event that has not yet been reflected in the objective data. Other experts may have information about planned changes and this might help in the selection of methods to forecast trend.

Rule-based forecasting explains why the methods are selected and combined. This thinking is fully exposed to the user. Improvements can be made by improving the rules for that situation and also by gaining more information

about the situation. We believe that this procedure will lead to improvements in the selection and combination of forecasts.

Our initial results suggest that rule-based forecasting provides a more effective selection: Combinations from a small set of methods were more accurate than those based on previous selection methods in tests using ex ante forecasts (Collopy and Armstrong, 1989).

Research Areas

Our ability to provide specific guidance about combining is limited in at least four key areas. These areas can be summarized by four questions: First, under what conditions is combining most useful? Second, what methods should be used to select the component forecasts? Third, is the simple average the best way to combine forecasts in all situations? Fourth, to what extent can accuracy be increased by combining forecasts? Further research is needed in these areas.

Under what conditions are combined forecasts most effective?

Practicing forecasters must make decisions about when to use combined forecasts. For example, is combining more useful for long-range or for short-range forecasting? One argument says that combining is more useful for long-range forecasting because of the greater uncertainty. An alternative viewpoint is that random errors are more significant for short-range forecasts; because these errors are off-setting, a combined forecast should reduce the errors. Initial results from my meta-analysis favored the latter view.

Rule-based forecasting is designed to specify the conditions under which combining is most useful. The rules can be easily revised as new knowledge is gained about the conditions.

What methods should be used to provide the forecasts?

Forecasters are advised to select a set of methods that differ substantially from one another with respect to the data used and also with respect to the procedures for analyzing the data (e.g., extrapolation or econometric or judgmental). The research offers little guidance on how to implement this advice.

One approach is to try a variety of feasible methods and then to simply combine the forecasts from each (Makridakis and Winkler, 1983). An alternative is to try to identify which methods are most appropriate and to use only them. This latter approach has not proven useful to date. However, the use of realistic simulations should lead to gains. We also expect rule-based forecasting to be helpful in this area.

Is the simple average best for all situations?

As noted, the use of a simple average has proven to do as well as more sophisticated approaches. An alternative simple approach, the median, might offer additional benefits. It is less likely to be affected by errors in the data. Whether the median is superior to the mean is an empirical issue. Meta-analysis may prove useful here. Two studies that address this issue (Agnew, 1985; Larréché and Moinpour, 1983) suggest that the median would improve accuracy.

Certainly, there are situations where one method is more accurate than another. If such cases can be identified in advance, simple averages would be inefficient. Realistic simulations of the actual forecasting situation could provide estimates of the relative accuracy of each method. These measures of relative accuracy could then be used in weighing the various forecasts. In addition, measures of dissimilarity would also be used; that is, the gains are only possible if the methods provide forecasts that differ.

Alternatively, rule-based forecasting might be used to weigh the different methods. This would draw upon prior evidence on which methods have been most accurate in a given type of situation. In addition, it could incorporate information about the specific situation.

Degree of improvement in accuracy

Combining increases costs. It is worthwhile then to examine the benefits of combining. To what extent does combining reduce the forecast error? An important study of this issue was conducted by Makridakis and Winkler (1983) in their examination of data from 1001 data series. One finding was that the combination of two methods led to a 7 % reduction in MAPE.

As suggested above, the combination of dissimilar methods would be expected to be more useful than combining similar methods. In Armstrong (1986), I conducted a meta-analysis to estimate the error reduction when different types of forecasting methods were used to prepare the component forecasts. Only eight studies were found. In these, a combination of two forecasts reduced the MAPE by almost 7% . Surprisingly, this was no better than the Makridakis and Winkler (1983) estimate for combining two similar methods.

Conclusions

Clemen's review tells us where we are now and it provides practical advice: combine and average. It also reveals shortcomings in our knowledge. Little specific guidance is available about the conditions under which combining is most effective, what methods should be combined, how they should be combined, and what gains in accuracy might be expected given different situations and different procedures.

Meta-analyses may offer some help in addressing the key questions. However, initial attempts have proven to be more difficult than anticipated because the criteria vary and the conditions are not well specified.

Realistic simulations should lead to improved selection and combining of forecasts. This is likely to refute the conclusion that simple averages are optimal. This approach does not, however, explain why the various methods work well.

Rule-based forecasting should help in the selection and combination of forecasts. In this procedure, the combination of *elements* of forecasts should lead to a better understanding as well as more accuracy. Initial results suggest that rule-based forecasts are more accurate than combined forecasts.

Realistic simulations and rule-based forecasting offer the promise of a new beginning for research on combined forecasts.

References

- Agnew, Carson E. (1985), "Bayesian consensus forecasts of macroeconomic variables," *Journal of Forecasting*, 4, 363-376.
- Armstrong, J. Scott (1985), *Long-Range Forecasting*. New York: Wiley.
- Armstrong, J. Scott (1986), "Research on forecasting: A quarter-century review, 1960-1984," *Interfaces*, 16, 89-109.
- Clemen, Robert T. (1989), "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559-583.
- Collopy, Fred and J. Scott Armstrong, (1992), "Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations," *Management Science*, 38, 1394-1414 (originally cited as a 1989 working paper).

Larréché, Jean-Claude and R. Moinpour (1983), "Managerial judgment in marketing: The concept of expertise," *Journal of Marketing Research*, 20,110-121.

Makridakis, Spyros and Robert L. Winkler (1983), "Averages of forecasts: Some empirical results," *Management Science*, 29, 987-996.