Forecast combination and encompassing: Reconciling two divergent literatures *

Francis X. DIEBOLD

Department of Economics, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract: The relationship of forecast combination to recent developments in the theory and practice of model selection (in particular, the encompassing literature) is explored, the pragmatic virtues of forecast combination are argued, and a synthesis is attempted. Promising areas for future research, particularly exploration and development of forecast combination methodologies that facilitate 'shrinkage' of estimated combining weights toward a measure of central tendency, are also discussed.

Keywords: Forecasting, Prediction, Pooling, Model selection, Shrinkage estimation.

1. Introduction

I believe that research in forecast combination, so nicely catalogued by Clemen (1989) in his comprehensive bibliography, has been a worthwhile and productive undertaking. Many econometricians disagree. In section 2, I examine this divergence of opinion and argue that it essentially stems from different loss functions. The relationship of forecast combination to recent developments in the theory and practice of model selection (in particular, the encompassing literature) is explored, the pragmatic virtues of forecast combination are argued, and a synthesis is attempted. Qualification is provided, however, regarding what I perceive to be an abuse – advocated by Clemen – of the forecast combination methodology. Next, in section 3, I highlight aspects of a promising area for future research: exploration and development of forecast combination methodologies that facilitate 'shrinkage' of estimated combining weights toward a measure of central tendency,

* This paper is a discussion of "Combining forecasts: A review and annotated bibliography", by Robert T. Clemen, which was presented at the 1989 International Symposium on Forecasting and is now published in the *International Journal of Forecasting*, this issue, pp. 559–583. I would like to thank Neil Ericsson for useful comments on an earlier draft.

such as the arithmetic average. I conclude in section 4 with a discussion of the need for evaluation of combined forecasts in truly ex ante, real-time environments.

2. Reconciling two divergent literatures

2.1. The econometric paradigm and the combination-of-forecasts paradigm

To understand the structural relationships in an economic system is to be able to forecast it, but not conversely. For this reason and others, many academic economists and econometricians tend to view the forecasting problem as one of secondary importance, claiming primary interest in understanding the economy, resting secure in the knowledge that good (or at least the best obtainable) forecasts will follow automatically from such understanding. There is a long and unfortunate tradition in economics, however, of placing too much emphasis on theory – as opposed to evidence (data) – in guiding the specification and evaluation of forecasting models. Within such a framework, one is led to focus on the task of estimating the parameters of a relationship suggested by a priori theory (at the cost of neglecting the model selection problem), after which the optimal prediction problem is easily solved (condi-

tional upon the assumed model). There is no role for forecast combination within such a paradigm.

Econometrics has changed drastically in the last ten or fifteen years; heavy emphasis is now placed on the use of data as well as on theory in model specification and evaluation. We now have available a large and useful set of tests for serial correlation, heteroscedasticity, adequacy of functional form, exogeneity, disturbance normality, parameter constancy, invalid restrictions and simultaneity bias. Moreover, the useful and complementary diagnostic role of tests of non-nested hypotheses has recently come to the fore. [1] In short, we now actively acknowledge the possibility of misspecification, and our focus has changed from effectively assuming the form of the true model known to one of cumulatively finding truth, or less idealistically, finding the model that provides the best explanation of the data. One requirement of such a model is that it be capable of explaining the results obtained from rival models, a phenomenon dubbed 'encompassing' by Hendry and his co-authors.

Again, however, there would appear to be little room for forecast combination within the encompassing paradigm. If a particular model appears misspecified, it is modified until an acceptable specification has been found, from which optimal forecasts follow directly. Forecast combination is readily dismissed; mechanical combination of manifestly incorrect models will never lead us to truth. Instead, it is argued, time is better spent improving specification. Chong and Hendry (1986) summarize this view:

> "...the composite artificial model which might be considered for forecast encompassing essentially coincides with the 'pooling of forecasts' formula...Note that the *need* to pool forecasts is *prima facie* evidence of a failure to encompass, and if H_1 is an econometric model and H_2 a univariate time series model (say) then if H_1 does not encompass H_2 it seems highly suggestive of the possibility that H_1 is dynamically misspecified..." (p. 677)

Similarly, Ericsson (1989) relegates forecast combination to a footnote, noting only that:

> "An extensive literature has developed on the pooling of forecasts, i.e., taking some (usually linear) combination of forecasts from different models to obtain a new forecast which has better properties, usually being a smaller MSFE than any of the individual model forecasts. Given the discussion above, finding such a pooled forecast is *prima facie* evidence of all individual models being mis-specified and may well indicate that a single model can be constructed which has smaller MSFE than even the pooled forecast." (p. 12)

This viewpoint leads one to think of combining regressions merely as specification (in particular, encompassing) tests, as originally advocated in Nelson (1972), Cooper and Nelson (1975) and Nelson (1984), and substantially extended by Chong and Hendry (1986).

2.2. How to use forecast combination

What, then, is the proper role of forecast combination techniques? In a world in which information sets can be instantaneously and costlessly combined, there is no role; it is always optimal to combine information sets rather than forecasts. The encompassing methodology then delivers the goods: it is explicitly designed to facilitate the combination of information sets.

However, while pooling of forecasts is suboptimal relative to pooling of information sets, it is clear that in many forecasting situations, particularly when deadlines must be met and timely forecasts produced, pooling of information sets is either impossible or prohibitively costly. This pragmatic observation is, entirely, the raison d'être of the forecast combination literature, in which forecasts rather than models are the basic object of analysis, reflecting an assumed inability to combine information sets.

The contrast is clear. The econometric approach takes a longer-run perspective, encouraging cumulative scientific progress toward a model that captures the salient features of the data-generating process and encompasses its rivals. The pooling approach, on the other hand, takes a distinctly short-run perspective, focusing on how to best patch things up along the rocky road to The Long Run. It should also be clear, however, that the two approaches are highly complementary, and that forecast combination can be viewed as a key link between the short-run, real-time forecast production process, and the longer-run,

---

[1] See Cox (1961, 1962) and Pesaran (1974), as extended by Hendry and Richard (1982), Mizon and Richard (1986), Chong and Hendry (1986) and Ericsson (1989).

ongoing process of model development. Perhaps the potential for misuse of forecast combination techniques, to which I now turn, is responsible for econometricians' reluctance to concur.

2.3. How not to use forecast combination

Mae West once said: "Too much of a good thing is wonderful." Such is not true of forecast combination. In particular, it should be stressed that if time and resource constraints do not prohibit the combination of information sets, then information sets, not forecasts, should be combined. It is all too easy to take a mechanical view of pooling ("I'll build five models and pool their forecasts"), which promotes taking forecast combination as an end in itself, rather than as a part of a larger modeling strategy.

For this reason, I find the recommendations in Clemen's final paragraph hard to swallow. It is certainly true that "with the advent of inexpensive forecasting software for personal computers, virtually any decision maker can generate multiple forecasts of a time series", but it does not follow that mechanical combination is desirable. The reason is simple: when the *user* of the forecast is in fact the model *builder*, the possibilities for combination of information sets – as opposed to forecasts – are greatly enhanced. In short, it seems to me that the end result of " an automated forecasting system that generates a minimum-variance composite forecast from Box–Jenkins and Holt–Winters models" is little more than a sorry excuse for serious analysis, and is precisely what we do *not* want.

3. Optimal composites or simple averages?

It is well known that, if the underlying variances and covariances are known, then an optimal (in the sense of Granger and Ramanathan) pooled forecast must have mean-squared-prediction-error (MPSE) less than or equal to that of the best primary forecast. In finite samples of the size typically available, however, sampling error contaminates the combining weight estimates and renders the above result invalid. Furthermore, the problem of sampling error is exacerbated by the collinearity that typically exists among primary forecasts. Thus, while we hope to reduce out-of-

sample MSPE by combining, there is no guarantee.

An intimately related issue in forecast combination, with immediate practical implications, is the frequently found superiority of simple averages, relative to 'optimal' composites. Various explanations have been advanced, such as Kang's (1986) focus on multicollinearity, but none is completely satisfactory. Obviously, the imposition of an 'equal weights' constraint beneficially reduces (eliminates) variation in the estimated weights at the cost of possibly introducing bias; the evidence indicates that the benefits of imposing the equal weights constraint often exceed the costs, i.e., the net effect is an MSPE decrease. What is striking is that such an extreme informational imposition – one in which the data at hand are completely ignored – should frequently appear optimal!

Diebold and Pauly (1987), building on Clemen and Winkler (1986), show how this finding may be exploited by using Bayesian shrinkage techniques to allow for the incorporation of varying degrees of prior information into the estimation of combining weights; least-squares weights and the 'prior' weights then emerge as polar cases for the posterior mean. The actual posterior mean combining weights are a matrix weighted average of those for the two polar cases, with the exact location depending on prior precision, which can be estimated from the data using an empirical Bayes procedure. Such procedures, with shrinkage toward a measure of central tendency (e.g., the arithmetic mean), should play an increasing role in forecast combination. In this way the combining weights are coaxed toward the artihmetic mean, but the data are still allowed to speak, when (and if) they have something to say.

4. Concluding remarks

I believe that methodology is currently ahead of application in the forecast combination literature. We have classical variance–covariance pooling, regression-based pooling, time-varying weights, robust and latent-root weight estimation, serial correlation, cointegration and error-correction in combining regressions, and so on. Most of these contributions represent adaptations of well-known statistical and econometric techniques to the problem of estimating combining weights, and

many such methodological papers present only a short 'illustrative example'. A disturbingly large number of these show only in-sample results, which convey no information regarding the usefulness of a proposed method in out-of-sample forecasting. Others illustrate new techniques out-of-sample but on artificially generated data, which does provide an indication of the potential usefulness of a new tool – but only under the controlled conditions of the laboratory. Others reserve a holdout sample on which to evaluate the performance of combined forecasts with weights estimated from earlier data; this approach gives a better indication of out-of-sample forecasting performance on real data, but one is always left wondering how much searching was involved in the selection of the time series, sample period and the in-sample/out-of-sample split used for the example.

In short, my feeling is that we do not yet know nearly as much as it might appear about the efficacy of forecast combination. We need to sort through the plethora of currently-available combining techniques by ex ante, real-time forecast evaluations, using large and comprehensive datasets, such as that of Makridakis et al. (1983). This unglamorous but highly valuable exercise would facilitate at least a partial separation of wheat from chaff.

References

Chong, Y.Y. and D.F. Hendry, 1986, "Econometric evaluation of linear macroeconomic models", *Review of Economic Studies*, 53, 671–690.
Clemen, Robert, T., 1989, "Combining forecasts: A review and annotated bibliography", *International Journal of Forecasting*, 5, 559–583 (*this issue*).
Clemen, Robert, T. and R.L. Winkler, 1986, "Combining economic forecasts", *Journal of Business and Economic Statistics*, 4, 39–46.
Cooper, J.P. and C.R. Nelson, 1975, "The *ex-ante* prediction performance of the St-Louis and FRB-MIT-PENN econometric models, and some results on composite predictions", *Journal of Money, Credit and Banking*, 7, 1–32.
Cox, D.R., 1961, "Tests of separate families of hypotheses", in: J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Vol. I* (University of California Press, Berkeley, CA) 105–123.
Cox, D.R., 1962, "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society Series B*, 24, 406–424.
Diebold, F.X. and P. Pauly, 1990, "The use of prior information in forecast combination", *International Journal of Forecasting*, forthcoming.
Ericsson, N.R., 1989, "Parameter constancy, mean squared forecasts errors, and measuring forecast performance: An exposition and extensions", Unpublished manuscript, Federal Reserve Board, Washington, D.C.
Granger, C.W.J. and R. Ramanathan, 1984, "Improved methods of forecasting", *Journal of Forecasting*, 3, 197–204.
Hendry, D.F. and J.F. Richard, 1982, "On the formulation of empirical models in dynamic econometrics", *Journal of Econometrics*, 20, 3–33.
Kang, H., 1986, "Unstable weights in the combination of forecasts", *Management Science*, 32, 683–695.
Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, 1983, *The Forecasting Accuracy of Major Time Series Methods* (Wiley, London, UK).
Mizon, G.E. and J.F. Richard, 1986, "The encompassing principle and its application to testing non-nested hypotheses", *Econometrica*, 54, 657–678.
Nelson, C.R., 1972, "The prediction performance of the FRB-MIT-PENN model of the US economy", *American Economic Review*, 62, 902–917.
Nelson, C.R., 1984, "A benchmark for the accuracy of econometric forecasts of GNP", *Business Economics*, 19, 52–58.
Pesaran, M.H., 1974, "On the general problem of model selection", *Review of Economic Studies*, 41, 153–171.

Biography: F.X. DIEBOLD received his Ph.D. in Economics from the University of Pennsylvania in 1986. From 1986–1989 he worked as a research economist at the Board of Governors of the Federal Reserve System in Washington D.C., after which he returned to the University of Pennsylvania. He is holder of the Joseph A. Cohen Chair. His publications are in the areas of econometrics, forecasting, macroeconomics, and international economics and finance.