Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting 15*, 495–508.
Chatfield, C. (2000). *Time-series forecasting*, Chapman & Hall/CRC Press, Boca Raton.
Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time-series analysis and forecasting. *International Statistical Review 63*, 289–308.

Simple versus complex methods

Robert T. Clemen
Duke University, Durham, NC, USA

Congratulations (yet again) to Makridakis and Hibon for this most recent contribution to the study of the empirical performance of forecasting methods. The M3-Competition extends the previous work begun by Makridakis and Hibon (1979) and continued in the M- and M2-Competitions by including a broader array of forecasting methods, including neural networks and expert systems, and considering 3003 time series. The basic message remains the same as in the previous work: Simpler forecasting methods perform well on average. Statistically sophisticated methods tend to perform relatively poorly on average. The implication of the work appears to be that research and teaching of sophisticated statistical forecasting techniques is misguided. Instead, given the empirical evidence from the competitions, would it not be better to focus on the development of more accurate methods?

The message of the M-Competitions is frustrating to those whose work focuses on the development of new and improved statistical forecasting techniques. In my case, I studied forecast combination methods. Although it is true that combining forecasts generally improves accuracy, it is also the case that simpler combining methods tend to perform well relative to more complex ones (Clemen, 1989). This basic message is once again reinforced by the results of the M3-Competition.

Why do more sophisticated methods perform relatively poorly? Received wisdom, stated again by Makridakis and Hibon in the present paper, is that real-world economic time series tend not to be stationary, at least not within the time frames for which forecasts are calculated. More sophisticated methods typically require estimation of more parameters, for which longer periods of stationarity are crucial.

Should the results of the M-Competitions lead us to forego teaching and research on the more sophisticated models? Clearly not. For one thing, economic applications represent only one arena in which forecasting models are used. Natural science applications abound, and in those cases sophisticated methods can perform well. Also, even in business and economic applications, there is a big difference between finding a method that works well for a single, important series and finding one that performs well on average over 1001 or 3003 time series. An analyst studying a particular series for his or her firm will do well to understand that series in as much depth as possible. Studying the series in the context of ARIMA models can provide considerable insight, even if the specification of the ARIMA model defaults to a relatively simple forecasting model, such as some form of exponential smoothing. Moreover, an analyst might consider a regression model with covariates as well as lagged dependent variables; doing a good job of specifying and estimating such a 'sophisticated' model requires considerable statistical skill and insight.

All these arguments notwithstanding, the results of all of the M-Competitions, culminating in the M3-Competition reported here, should give all forecasting researchers pause. If our goal is to develop methods for use in many situations, it is indeed time to understand why simple methods perform well relative to more complex ones. If the reason really is nonstationarity, then the challenge is to develop robust ways to forecast within that nonstationary environment.

References

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography (with discussion). *International Journal of Forecasting 5*, 559–583.

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, Series A 142*(Part 2), 97–145.

Explaining the results of the M3 forecasting competition

Michael P. Clements[a,*], David F. Hendry[b,1]

[a]*Warwick University, Coventry, UK*
[b]*Oxford University, Oxford, UK*

Makridakis and Hibon (2000) summarize four main implications of the latest forecasting competition, which we paraphrase as: (a) 'simple methods do best'; (b) 'the accuracy measure matters'; (c) 'pooling helps'; and (d) 'the evaluation horizon matters'. We applaud the detailed empirical investigations, are unsurprised by their summary; but are surprised by the assertion that 'the strong empirical evidence, however, has been ignored by theoretical statisticians'. Having successfully published two books and more than a dozen papers across a wide range of journals, which *inter alia* analyze their four points, we refute the claim that the issue is being 'ignored', and doubt the implicit suggestion of hostility by the profession.

What must be the relationship between the world to be forecast and the models with which we forecast for conditions (a)–(d) not to hold? The research summarized in Clements and Hendry (1998b, 1999) (henceforth CH98 and CH99) shows that in weakly stationary processes, a congruent, encompassing model in-

*Corresponding author.

[1]Financial support from the U.K. Economic and Social Research Council under grant L116251015 is gratefully acknowledged by both authors.