Forecasting by Extrapolation: Conclusions from Twenty-five Years of Research

J. Scott Armstrong

University of Pennsylvania, armstrong@wharton.upenn.edu

Forecasting by Extrapolation: Conclusions from 25 Years of Research

J. Scott Armstrong
Wharton School, University of Pennsylvania

Sophisticated extrapolation techniques have had a negligible payoff for accuracy in forecasting. As a result, major changes are proposed for the allocation of the funds for- future research on extrapolation. Meanwhile, simple methods and the combination of forecasts are recommended.

Have advances in extrapolation methods helped to make short-range forecasts better now than in 1960? I am defining extrapolation as methods that rely solely on historical data from the series to be forecast. No other information is used. This class of methods is widely used in forecasting, especially for inventory control, process control, and in situations where other relevant data are not available.

I describe a forecasting procedure that was used in 1960 and then present evidence from research published over the last quarter of a century.

Short-Range Forecasting in 1960

As an industrial engineer at Eastman Kodak in the early 1960s, I examined the short-range forecasting system for color print orders from customers. These forecasts were needed to schedule part-time workers and to control inventories. The procedure that had been used for many years prior to 1960 had the following characteristics:

- Weekly historical data on each type of order were collected on the basis of billing records. In general, these data were thought to be accurate. Outliers were adjusted or removed.

- Graphs were prepared for the more important items.

- The forecasts were then prepared judgmentally by a man who had been doing this job for many years.

Existing literature implied that better forecasts could be obtained by using objective methods. Accordingly, I developed and tested a model that used exponential smoothing of deseasonalized data. The deseasonalizing also included adjustments for trading days and holidays. Search procedures were then used to find the most appropriate smoothing factors for the average and the additive trend. The procedures were based primarily on Brown [1959] and Shiskin [1958].

Historical comparisons showed that the exponential smoothing model was superior to the judgmental method for almost all items. Side-by-side comparisons over the next six months provided additional evidence on the superiority of the exponential smoothing model.

Given what we now know, would it have been possible to improve upon this extrapolation model, which had been developed with methods described in publications prior to 1960?

Progress Since 1960

Many sophisticated approaches to extrapolation methods have been developed since 1960. Sophistication has come in two forms: sophisticated methods are used to select the appropriate type of model (for example, Box-Jenkins [1976] procedures), and complex models are used to analyze historical data.

This sophistication calls for a fitter understanding of mathematics. It has also led to an increase in jargon. (For example, the term "univariate time series methods" is often used instead of "extrapolation.") Sophisticated approaches frequently lead to more complex forecasting models which are difficult to understand. Forecasters tell me that few of the people in their organizations understand these sophisticated methods.

Numerous theoretical papers have made claims for the superiority of the more sophisticated methods. Often, these methods are shown to provide a better fit to historical data.

Are these more sophisticated methods better? One can examine this question using many criteria. On the *cost* side, more sophisticated methods are generally more difficult to understand, and they cost more to develop, maintain, and operate. On the *benefit* side, more sophisticated methods may be expected to produce more accurate forecasts and to provide a better assessment of uncertainty. (Other criteria are also important, but in this paper I am considering only those related to forecasting.)

An assessment of the relative accuracy of sophisticated versus simple methods is provided in Table 1. This table lists all the published studies I could find (although I would not claim that the list is exhaustive). In addition to a library search, I relied heavily on experts in the field to identify relevant studies. Gardner [1985] was a particularly helpful source. Only published empirical studies that provided forecast comparisons were included. My ratings of sophistication were subjective and others might have different opinions. The simple end of the continuum includes "no change" or "constant trend" models. Exponential smoothing was regarded as more sophisticated than moving averages, despite the fact that exponential smoothing is less expensive. Added sophistication is provided by methods where the parameters change as new observations are received, or by methods that use more terms or more complex relationships to time. Box-Jenkins procedures provide a sophisticated approach to selecting and estimating the proper model.

Table 1 shows that the scientific evaluation of extrapolation methods is comparatively recent. None of the 39 studies were published prior to 1960. Six studies were published in the 1960s, and 25 were published in the 1970s. This indicates a rapid growth rate in evaluation research on extrapolation methods.

More important, Table 1 provides little evidence to suggest that sophistication beyond the methods available in the 1960s has had any payoff. Relatively simple methods seem to offer comparable accuracy; 21 studies concluded that there were negligible differences, and for the 18 studies showing differences, 11 favored sophistication, and seven favored simplicity. However, of the 11 cases favoring sophistication, three have since been challenged, and three cases were based on the superiority of exponential smoothing (available prior to 1960). We are left with five comparisons favoring sophistication and seven favoring simplicity.

In general, the findings on sophisticated methods are puzzling, and it is unlikely that they could have been anticipated in 1960. Many of the sophisticated methods made good sense. For example, it seemed reasonable to expect that models in which the parameters are automatically revised as new information comes in should be more accurate. Results of the studies offer little support to this viewpoint: Four studies favor the use of adaptive parameters, three (including also Torfin and Hoffman [1968]) show no difference, and three suggest that adaptive parameters are less accurate. Furthermore, Gardner's [1983] reevaluation of Chow's (1965) data yielded different conclusions, and Ekern's [1981] reanalyses of data from the Whybark [1973] and Dennis [1978] studies challenged their conclusions, so three of the positive findings did not hold up. The box score for models with adaptive parameters does not look good so far. A simulation study by Kahl and Ledolter (1983) suggested that adaptive methods help if, in fact, the parameters do vary. Otherwise, adaptive methods react unfavorably to random noise and lead to greater errors. But Kahl and Ledolter questioned whether it is possible to identify situations where the parameters do change in the real world. It does suggest that manual intervention might help when other evidence clearly indicates that a change has occurred (for example, the new tax laws might affect the relationships).

Highly complex models may reduce accuracy. While these complex models provide better fits to historical data, this superiority does not hold for forecasting. The danger is especially serious when limited historical data are available. Ledolter and Abraham [1981] show that unnecessary parameters in the estimation model will increase the mean square forecast error by $1/n$ (where n is the number of historical observations). Bunn [1980], in a simulation study, found that the error was larger when the model was too complex for the situation. Conversely, he found larger errors if the model was too simple (for example, no trend term in situations where a trend existed).

Of particular interest in the evaluations listed in Table 1 are the large scale empirical studies by Makridakis and Hibon [1979], which examined forecasts for 111 time series, and Makridakis et al. [1982], which examined forecasts for 1001 time series (the latter study being known as the M-Competition). In general, these studies did not offer much support for sophisticated methods. Both studies were subjected to much re-examination, especially the M-Competition (see Armstrong and Lusk [19831]. McLaughlin [1983], one of the commentators in the M-Competition, reanalyzed forecasts from 15 of the methods, and found that a naive model based on the assumption that the next period would be the same as the last (using seasonally adjusted data) was superior to more sophisticated extrapolations for nine of 15 comparisons.

Table 1
The accuracy of sophisticated versus simple methods for forecasting. For "results," a + means sophisticated methods were more accurate, a 0 means negligible difference, and a – means simple methods were more accurate.

Study	Major Comparisons	Results
Winters [1960]	exponential smoothing versus moving averages	+
Frank [1969]	exponential smoothing versus moving averages	+
Elton and Gruber [1972]	exponential smoothing versus moving averages	+
Chow [1965]	adaptive versus constant parameters	+
Whybark [1972]	adaptive versus constant parameters	+
Smith [1974]	adaptive versus constant parameters	+
Dennis [1978]	adaptive versus constant parameters	+
Brown and Rozeff [1978]	Box-Jenkins versus simple trend	+
Newbold and Granger [1974]	Box-Jenkins versus exponential smoothing	+
Reid [1975]	Box-Jenkins versus exponential smoothing	+
Dalrymple [1978]	Box-Jenkins versus regression	+
Kirby [1966]	exponential smoothing versus moving averages	0
Adam [1973]	exponential smoothing versus moving averages	0
Raine [1971]	adaptive versus constant parameters	0
Dancer and Gray [1977]	adaptive versus constant parameters	0
Chatfield and Prothero [1973]	Box-Jenkins versus no-change	0
Albrecht et al. [1977]	Box-Jenkins versus no-change	0
Bates and Granger [1969]	Box-Jenkins versus exponential smoothing	0
Groff [1973]	Box-Jenkins versus exponential smoothing	0
Geurts and Ibrahim [1975]	Box-Jenkins versus exponential smoothing	0
Mabert [1976]	Box-Jenkins versus exponential smoothing	0
Chatfield [1978]	Box-Jenkins versus exponential smoothing	0
Kenny and Durbin [1982]	Box-Jenkins versus exponential smoothing	0
Torfin and Hoffman [1968]	6 models of varying complexity	0
Markland [1970]	4 models of varying complexity	0
Johnson and Schmitt [1974]	10 models of varying complexity	0
Carey [1978]	21 models of varying complexity	0
Hagerman and Ruland [1979]	3 models of varying complexity	0
Makridakis and Hibon [1979]	22 models of varying complexity	0
Ruland [1980]	8 models of varying complexity	0
Makridakis et al. [1982]	21 models of varying complexity	0
Armstrong [1975]	complex curve versus rule of thumb	0
Mabert [1978]	adaptive versus constant parameters	–
Gardner [1979]	25 models of varying complexity	–
Gardner and Dannenbring [1980]	adaptive versus constant parameters	–
McLeavey, Lee and Adam [1981]	adaptive versus constant parameters	–
Ledolter and Abraham [1981]	simple versus complex models	–
Coggin and Hunter [1982-3]	simple versus complex models	–
Brandon, Jarrett and Khumawala [1983]	Box-Jenkins versus 8 simple models	–

Some have suggested that sophisticated methods will prove more accurate if *properly used* by experts (for example, see Chatfield [1978, p. 266]). Is it important to match the sophistication of the method with the expertise of the user? Evidence to date suggests, surprisingly, that this concern is not an important one, given that the user has achieved a certain (seemingly modest) level of expertise. For example, the M-Competition results were largely supportive of Makridakis and Hibon [1979] even though the M-Competition was designed to have outstanding experts for each method. Automatic procedures produced accuracy from sophisticated methods equivalent to that produced when these methods were used by experts [Hill and Fildes 1984]; this suggests that there is little need for experts. Also, a study by Carbone, Andersen, Corriveau, and Corson [1983] found that students with a modest amount of instruction were able to do as well as experts when using Box-Jenkins and other methods.

Some have suggested, based on *ex post* interpretations of the M- Competition, that sophistication might be relevant in certain situations. I would have found this argument more convincing had their hypotheses on "which methods would be better in what situations" been stated *prior* to the M-Competition.

My conclusions about the value of simple methods for forecasting in business and economics remind me of the history of extrapolation methods for demographic forecasting. Dorn [1950] and Hajnal [1955], in their reviews of the evidence, both suggested that the more thought demographers put into their extrapolations, the more complex their methods became, but the poorer their accuracy.

Alternatives to Sophistication

Although the development of sophisticated methods has yielded little gain to date, some areas do seem worthy of further research. It may be possible to clearly identify which methods will be better in given situations. To do this, I suggest experimental studies comparing specific methods, rather than the approach used in the M-Competition, where models that utilized a variety of methods were contrasted. For example, one might study whether the attenuation (or dampening) of trend factors would produce better forecasts in situations involving large changes. Also, one could examine whether the attenuation of seasonal factors improves accuracy in situations with large but difficult to measure seasonal factors.

Lawrence [1983] concluded that people can do as well as computers and sophisticated methods in extrapolating trends. These results were obtained even though the people had no information other than the historical data on the series being forecasted. In other words, the extrapolation by person was based on the same information as the extrapolation by computer. More recent results in Lawrence, Edmundson, and O'Connor [1985] support Lawrence's original findings. These studies suggest that further research is needed on when and how to use judgment for extrapolation.

Still another area of promise is the strategy of combining forecasts from different extrapolation methods. Newbold and Granger [1974], in a study of 80 monthly time series, showed that the best forecast can usually be improved by combining it with a forecast from another extrapolation method. Morris [1977], in a forecast of the sunspot cycle, reduced the error by 50 percent when combining forecasts from two different extrapolation models; however, the weights on each forecast were selected after the fact, so the gain is suspect. Additional evidence is provided in Table 2, which summarizes all studies that I could find that made an estimate of the magnitude of error reduction that resulted from combining forecasts. The Bates and Granger and the Ogburn results are based on my reanalysis of their results. The combinations produced significant gains in comparison with the typical (average) forecast error of the methods used, but the sample sizes for these studies were small.

Table 2
Percentage reduction in error by combining extrapolation forecasts versus average error of component forecasts

Study	Data	Criteria*	Number or Methods			
			2	3	4	5
Armstrong [1978, p. 167]	International Photographic Market	MAPE	17	37	-	-
Bates and Granger [1969]	International Air Travel	RMSE	12	-	-	-
Ogburn [1946]	Air Travel Market	MAPE	-	-	-	64
Reinmuth and Geurts [1979]	Salt Lake Retail Sales	Theil's U2	80	92	-	-
Bunn [1979]	Tourists to Hawaii	RMSE	60	-	-	-

* MAPE = Mean Absolute Percentage Error; RMSE = Root Mean Square Error

Additional and much more extensive evidence on the value of combined forecasts was presented by Makridakis and Winkler [1983] using data from the 1001 Series in the M-Competition data. If one does not know which of a set of extrapolation forecasts is most accurate (this being the typical case), then combining forecasts produces dramatic gains. Although the extent of the gains diminished with each additional method in their study, substantial benefits were made even for the fourth or fifth methods.

The question of how many forecasts to combine is, of course, a cost/benefit issue. Even if you know which is the best extrapolation method (an unusual situation), the strategy of combining forecasts, is likely to help, especially if a number methods are combined.

Given that we know little about the best way to weight the components of a combined forecast, I suggest starting with the least expensive method, and then investing in successively more expensive methods. Or you could start with the most understandable method. Use methods that are as different as possible, and simply weight each forecast equally.

Conspicuous by its absence is research on the best way to assess uncertainty in these forecasts. Williams and Goodman [1971] and Newbold and Granger [1974] did find that the confidence intervals developed by simulating the use of the extrapolation.

Conclusions

Since 1960, significant effort has been devoted to the development of sophisticated methods for forecasting by extrapolation. Shortly after this, research began to evaluate these new methods. Research on the evaluation of extrapolation methods has been growing rapidly, and evidence is now available from about 40 studies.

For further research on extrapolation, experimental designs should be used to assess hypotheses on each component in specific situations (for example, what is the value of attenuating the trend where large changes are forecasted?), the assessment of the most effective way to use experts, the use of combined forecasts, and the best way to assess uncertainty.

In addition to using empirical and simulated data, researchers can test hypotheses using previously published research. Only recently has the number of studies become large enough to allow for this type of research. *Quantitative* reviews of this research, "meta-analysis," would be especially valuable as demonstrated in the experiment by Cooper and Rosenthal [1980].

For the practitioner, the implications are clear: Relatively simple extrapolation methods like those described by Brown [1959], Shiskin [1958], and Winters [1960] are adequate. The analyst should make adjustments for outliers, trading days, and seasonality, then use a relatively simple method such as exponential smoothing to estimate trend and seasonal factors. These methods provide accuracy equivalent to more complex methods at a lower cost, and they are easier to understand. Remember the advice of William of Occam, who said "one should not introduce complexities unless absolutely necessary." Despite the fact that William died of the Black Death in 1349, "Occam's Razor" is good advice. Finally, use more than one method. My suggestion is to spread the budget over three or four in expensive and easily understood. methods, then calculate an average forecast.

References

Adam, Everett, Jr. (1973), "Individual item forecasting model evaluation," *Decision Sciences,* 4 (4), 458-470.

Albrecht, W. S., Lookabill, L. L. and McKeown, J. C. (1977), "The time series properties of annual earnings," *Journal of Accounting Research*, 15 (2), 226-244.

Armstrong, J. Scott (1975), "Monetary incentives in mail surveys," *Public Opinion Quarterly*, 39 (1), 111-116.

Armstrong, J. Scott (1978)*, Long-Range Forecasting: From Crystal Ball to Computer*. New York: John Wiley.

Armstrong, J. Scott and Lusk, Edward J. (1983), "The accuracy of alternative extrapolation models: Analysis of a forecasting competition through open peer review," *Journal of Forecasting*, 2 (3), 259-262, with commentary and replies, 263-311.

Bates, J. M. and Granger, C. W. J. (1969), "The combination of forecasts," *Operational Research Quarterly*, 20 (4), 451-468.

Box, George E. and Jenkins, G. M. (1976)*, Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Brandon, C. H., Jarrett, J. E. and Khumawala, S. B. (1983), "Revising forecasts of accounting earnings: A comparison with the Box-Jenkins method," *Management Science*, 29 (2), 256-263.

Brown, L. D. and Rozeff, M. S. (1978), "The superiority of analyst forecasts as measures of expectations: Evidence from earnings," *Journal of Finance*, 33 (1), 1-16.

Brown, Robert G. (1959), *Statistical Forecasting for Inventory Control.* New York: McGraw-Hill.

Bunn, Derek W. (1979), "The synthesis of predictive models in marketing research," *Journal of Marketing Research,* 16 (2), 280-283.

Carbone, Robert, Andersen, A., Corriveau, J. and Corson, P. P. (1983), "Comparing for different time series methods the value of technical expertise, individualized analysis and judgmental adjustment," *Management Science,* 29 (5), 559-566.

Carey, K. J. (1978), "The accuracy of estimates of earnings from naive models," *Journal of Economics and Business,* 30 (3), 182-193.

Chatfield, C.(1978), "The Holt-Winters forecasting procedure," *Applied Statistics*, 27 (3), 64-279.

Chatfield, C. and Prothero, D. L. (1973), "Box Jenkins seasonal forecasting: Problems in a case study," *Journal of the Royal Statistical Society*: Series A, 136 (3), 295-352.

Chow, W. M. (1965), "Adaptive control of the exponential smoothing constant," *Journal of Industrial Engineering,* 16 (5), 314-317.

Coggin, T. Daniel and Huntor, John E. (1982-3), "Analysts EPS forecasts nearer actual than statistical models," *Journal of Business Forecasting*, 1 (6), 20-23.

Cooper, Harris M. and Rosenthal, Robert (1980), "Statistical versus traditional procedures for summarizing research findings," *Psychological Bulletin*, 87 (3), 442-449.

Dalrymple, Douglas J. (1978), "Using Box Jenkins in sales forecasting," *Journal of Business Research,* 6 (2), 133-145.

Dancer, Robert and Gray, Clifford (1977), "An empirical evaluation of constant and adaptive computer forecasting models for inventory control," *Decision Sciences*, 8 (1), 228-238.

Dennis, J. D. (1978), "A performance test of a run-based adaptive exponential forecasting technique," *Production and Inventory Management*, 19 (2), 43-46.

Dorm Harold F. (1950), "Pitfalls in population forecasts and projections," *Journal of the American Statistical Association*, 45 (251), 311-334.

Ekern, Steinar (1981), "Adaptive exponential smoothing revisited," *Journal of the Operational Research Society*, 32 (9), 775- 782.

Elton, Edwin J. and Gruber, Martin J. 1972, "Earnings estimates and the accuracy of expectational data," *Management Science,* 18 (8), B409-B424.

Frank, Werner (1969), "A Study of the predictive significance of two income measures," *Journal of Accounting Research*, 7 (1),123-136.

Gardner, Everette S., Jr. (1979), "Box-Jenkins vs. multiple regression: Some adventures in forecasting the demand for blood tests," *Interfaces*, 9 (4), 49-54.

Gardner, Everette S., Jr. (1983), "Evolutionary operation of the exponential smoothing parameter: Revisited," *Omega,* 11 (6), 621-623.

Gardner, Everette S., Jr. (1985), "Exponential smoothing: The state of the art," *Journal of Forecasting*, 4 (1), 1-28. Commentary follows on 29-38.

Gardner, Everette S., Jr. and Dannenbring, David G. (1980), "Forecasting with exponential smoothing: Some guidelines for model selection," *Decision Sciences*, 11 (2), 370-383.

Geurts, Michael D. and Ibrahim, I. B. (1975), "Comparing the Box-Jenkins approach with the exponentially smoothed forecasting model application to Hawaii tourists," *Journal of Marketing Research,* 12 (2), 182-188.

Groff, Gene K. (1973), "Empirical comparison of models for short range forecasting," *Management Science,* 20 (1), 22-31.

Hagerman, R. L. and Ruland, W. (1979), "The accuracy of management forecasts and forecasts of single alternative models," *Journal of Economics and Business*, 31 (3), 172-179.

Hajnal, John (1955), "The prospects for population forecasts," *Journal of the American Statistical Association*, 50 (270), 309-327.

Hill, Gareth and Fildes, Robert (1984), "The accuracy of extrapolation methods: An automatic Box-Jenkins package (SIFT)," *Journal of Forecasting*, 3, 319-323..

Johnson, T. E. and Schmitt, T. G. (1974)**,** "Effectiveness of earnings per share forecasts," *Financial Management,* 3 (2), 64-72.

Kahl, Douglas R. and Ledolter, Johannes (1983), "A recursive Kalman filter forecasting approach," *Management Science*, 29 (11), 1325-1333.

Kenny, Peter B. and Durbin, James (1982), "Local trend estimation and seasonal adjustment of economic and social time series" (with discussion), *Journal of the Royal Statistical Society:* Series A, 145 (1), 1-41.

Kirby, Robert M.(1966), "A comparison of short and medium range statistical forecasting methods," *Management Science*, 13 (4), B202-B210.

Lawrence, M. J. (1983), "An exploration of some practical issues in the use of quantitative forecasting models," *Journal of Forecasting*, 2 (2), 169-179.

Lawrence, M. J., Edmundson, R. H. and O'Connor, M. J. (1985), "An examination of the accuracy of judgmental extrapolation of time series," *International Journal of Forecasting*, 1 (1), 25-35.

Leolter, Johannes and Abraham, Bovas (1981), "Parsimony and its importance in time series forecasting," *Technometrics*, 23 (4*),* 411-414.

Mabert, Vincent A. (1976), "Statistical versus sales force-executive opinion short range forecasts: A time series case study," *Decision Sciences*, 7 (2), 310-318*.*

Mabert, Vincent A. (1978), "Forecast modification based upon residual analysis: A case study of check volume estimation," *Decision Sciences*, 9 (2), 285-296.

Makridakis, Spyros, Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1 (2), 111-153*.*

Makridakis, Spyros and Hibon, Michele (1979), "Accuracy of forecasting: An empirical investigation" with discussion, *Journal of the Royal Statistical Society:* Series A, 142 (2), 97-145.

Makridakis, Spyros and Winkler, Robert L. (1983), "Averages of forecasts: Some empirical results," *Management Science*, 29 (9), 987-996*.*

Markland, Robert E. (1970), "A comparative . study of demand forecasting techniques for military helicopter spare parts," *Naval Research Logistics Quarterly*, 17 (1), 103-119.

McLeavey, Dennis W., Lee, T. S., and Adam, E. E. (1981), "An empirical evaluation of individual item forecasting models," *Decision Sciences*, 12 (4), 708-714.

Morris, M. J. (1977), "Forecasting the sunspot cycle," *Journal of the Royal Statistical Society: Series A,* 140 (4), 437-468.

Newbold, Paul and Granger, C. W. J. (1974), "Experience with forecasting univariate time series and the combination of forecasts," *Journal of the Royal Statistical Society: (Series A),* 137 (2), 131-165.

Ogburn, William F. (1934), "Studies in prediction and distortion of reality," *Social Forces*, 13 (2), 224-229.

Raine, Jesse E. (1971), "Self-adaptive forecasting reconsidered," *Decision Sciences,* 2 (2), 181-191.

Reid, David J. (1975), "A review of short-term projection techniques," in *Practical Aspects of Forecasting.* H. A. Gordon, ed., London: Operational Research Society.

Reinmuth, James E. and Geurts, Michael D. (1979), "A multideterministic approach to forecasting," in *Forecasting*. S. Makridakis and S. C. Wheelwright, eds., New York: North Holland.

Ruland, W. (1980), "On the choice of simple extrapolative model forecasts of annual earnings," *Financial Management,* 9 (2), 30-37.

Shiskin, Julius (1958), "Decomposition of economic time series," *Science*, 128 (3338), 1539-1546.

Smith, David E. (1974), "Adaptive response for exponential smoothing: Comparative system analysis," *Operational Research Quarterly*, 25 (2), 421-435.

Torfin, Gary P. and Hoffman, T. R. (1968), "Simulation tests of some forecasting techniques," *Production and Inventory Management,* 9 (2), 72-78.13-26.

Whybark, D. Clay (1972), "A comparison of the surprising conclusions cited by adaptive forecasting techniques," *Logistics and Transportation Review*, 8 (3), 13-26.

Williams, W. H. and Goodman, M. L. 1971, "A simple method for the construction of empirical confidence limits. for economic forecasts," *Journal of the American Statistical Association,* 66 (336), 752-754.

Winkler, Robert L. and Makridakis, Spyros (1983), "The combination of forecasts," *Journal of the Royal Statistical Society: Series A*, 146 (2), 150-157.

Winters, Peter R. (1960), "Forecasting sales by exponentially weighted moving averages," *Management Science*, 6 (3), 324-342.

A comment from Robert U. Ayres, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

The forecasting profession (like some others) will benefit more from the kind of insight exemplified by Professor Armstrong's review paper than from more of the rather academic studies he surveys. Perhaps I may be permitted a few speculative remarks of an even more Olympian character.

The main conclusions to be drawn from this body of research are that more sophisticated models do not consistently yield better results than simple ones or even "eyeball" extrapolations but that combination of various methods consistently do give better results than any single method. Now, it is fairly obvious that the best possible forecast will be obtained by whatever procedure most accurately *decodes* the time series data, that is, most fully utilizes the message contained in it while rejecting the "noise." Each of the statistical methods tested only recovers part of the coded information. In fact, the surprising conclusions cited by Armstrong can perhaps be explained on the hypothesis that even the "best" statistical method recovers only a fairly small part of the coded information in the time series. The fact that subjective "eyeball" extrapolations are as good as computer extrapolations, at least in some situations, would seem to substantiate this hypothesis. It is clear that in this case (as in numerous others) the human brain does not utilize the same computational (or analog) procedure as the computer. Although I claim no special expertise in cognitive psychology, it is virtually certain that human beings do much better at forecasting trends presented in graphic forms than they do at forecasting from tabular data. The likelihood is that the human brain does two things better than statistical models: First, it extracts and synthesizes a good deal of information on second and higher order derivatives that is largely neglected by statistical models. Second, the brain is probably better at discriminating random noise from cyclic variations. The fact that the algorithm for doing these things is unknown to cognitive psychology does not mean that plenty of higher order information is not potentially recoverable from time series data.

The bottom line seems to be that even the best single statistical technique yet invented for decoding time series data may not be particularly good in absolute terms. This is a fascinating conjecture, startlingly reminiscent of a theorem from information theory (by Claude Shannon?) to the effect that even the most efficient known code is less efficient in terms of channel capacity utilization than the *average* of all possible codes. If this is true,

9

statisticians are in no immediate danger of approaching asymptotic limits. Unfortunately, it is by no means obvious how to break through the present impasse.

A comment from Carl F. Christ, The Johns Hopkins University, Baltimore, Maryland.
.

Readers of Scott Armstrong's valuable survey of forecasting by extrapolation may be interested in a brief look at another approach to forecasting: econometric modeling. There is substantial literature on it, and several forecasters now use it on a regular basis.

In 1975 I summarized the performance of such models as follows: "Econometric models of the US economy have been developed to the point where forecasters who use them can forecast real and nominal GNP two or three quarters ahead with root mean square errors of less than one percent, and six quarters ahead with RMS errors of one to two percent. The best of them now usually do better than forecasters who do not use such models." [Christ 1975, p. 54]

The situation is roughly similar today. Forecasters who use such models can do rather well over short horizons (up to six quarters ahead), and the best ones outdo forecasters who don't use such models. Many firms pay to receive such forecasts, which indicates that they have value. However, for longer horizons the forecasting quality deteriorates. Furthermore, the models diverge substantially from each other for longer-run forecasts. Hence there is roam for further improvements.

A thorough review of several leading econometric models and their performance is in Klein and Burmeister [1976]. A careful comparison of forecasts made by several models is in Fromm and Klein [1976], which is the source of the data for graphs of forecasts errors presented in Christ [1975] . See also Zarnowitz [1979] for more recent evaluations of forecasts.

A promising union of econometric modeling and Box-Jenkins time series methods is proposed by Zellner and Palm [1974]. They suggest that time-series methods be used to obtain the forecasts of the variables that are not explained by the econometric model. And they explain how tests of the econometric model can be conducted with the aid of time series analysis.

An important critique of econometric methodology is offered by Lucas [1976]. It is related to the theory of rational expectations. This theory holds that some of the variables to be explained are private agents' current expectations of future values of other variables, and that private agents will form these expectations in a rational way, based on the information they have (or can obtain at some cost) about the private economy and about the policy rules being followed by the authorities. Lucas notes that in consequence, some of the parameters in conventional econometric models are functions of government policy rules, and will change if the policy rules are changed. This introduces an added complication into the forecasting process. Current research is directed at this problem.

References

Christ, Carl F. (1975), "Judging the performance of econometric models of the US economy," *International Economic Review*, 16 (1), 54-74. (Reprinted in Klein and Burmeister, 1976, 322-342.)

Fromm, Gary, and Klein, Lawrence R. (1976), "The NBER/NSF model comparison seminar: An analysis of results," *Annals of Economic and Social Measure,* 5 (1), 1-28. (Reprinted in Klein and Burmeister, 1976, 380-407.)

Klein, Lawrence R., and Burmeister, Edwin, eds. (1976), *Econometric Model Performance,* Philadelphia: University of Pennsylvania Press.

Lucas, Robert E., Jr. (1976), "Econometric policy evaluation: A critique," in *The Phillips Curve and Labor Markets.* Karl Brunner and Allan H. Meltzer, eds., Vol. 1 of the *Carnegie-Rochester Conference Series on Public Policy*, 19-46.

Zarnowitz, Victor (1979), "An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level," *Journal of Business*, 52 (1), 1-33.

Zellner, Arnold, and Palm, Franz (1974), "Time series and simultaneous equation econometric models," *Journal of Econometrics,* 2 (1), 17-54.

───────────────────────

A comment from J. Keith Ord, Division of Management Science, The Pennsylvania State University, University Park, Pennsylvania.

In an earlier paper [Armstrong 1978], Scott Armstrong challenged the assumption that econometric models can provide better forecasts, and he has now presented a similarly well-researched and thought-provoking survey of extrapolation techniques. He notes that not all the earlier readers of his manuscript are in complete agreement with his views, so further dissent will come as no surprise. The several points appear under individual headings for clarity of presentation (and ease of reply!)

The Box-Jenkins Approach

In their seminal work, Box and Jenkins [1970] developed a four-stage approach to time-series analysis: model *identification* (or selection) from a broad class of possibles, the *estimation* of unknown parameters, model *validation,* and finally, *forecasting.* Armstrong is quite clear in his paper that "Box-Jenkins" is an approach to forecasting and not a specific method, although others have fudged the distinction. Nevertheless, it is useful to look at the various stages of the Box-Jenkins paradigm.

Model Identification: Box and Jenkins suggest that 30 or more observations are needed before their correlogram methods provide a reliable guide to model selection; even more observations are required for monthly seasonal data. Since the majority of the series in the M-Competition [Makridakis et al. 1982] were shorter than this, we might expect that the correlograms often gave a less than clear indication. Further, short series tend to lead to the selection of simple models which, in turn, often reduce to standard forecasting techniques. The best known example of this is the first order moving average model for a (once) differenced time-series which leads directly to single exponential smoothing. In such circumstances, the scarcity of data may lead the "Box-Jenkins expert" to the same model as the "simplistic forecaster" and, in turn, to the sane forecasts. The Box-Jenkins approach can hardly be blamed for failing to deliver more than is possible from a scanty database.

Model Estimation: The improvements in estimation procedures have been considerable in the past 10 years, and many of these are now being used even for simple forecasts. For example, in the M-Competition, Makridakis used discounted least squares for the exponential smoothing methods in preference to the less efficient methods used in Makridakis and Hibon[1979]. As computing costs fall, the introduction of improved estimation procedures offers scope for better forecasting.

Model Validation: Tests for checking a model's validity can lead us towards more automated procedures of selection. I agree with Scott Armstrong that there is a need to more clearly identify "which methods will be better in given situations." To me, this suggests more work on validation and selection methods, not less.

Forecasting: The optimality of sophisticated forecasts rests upon two assumptions: that the model is correct and that the model structure does not change over time (particularly, in the future!). The earlier steps ensure that we model as best we can, but a change of structure may still blow the forecasts off course. If, indeed, time-series in business follow the Hobbesian View of life and are "nasty, short, and brutish," then perhaps simple methods do as well as any other *single* series technique. However, structural changes may themselves be predictable, suggesting that

11

intervention analysis [Box and Tiao 1975] or Bayesian forecasting methods [Harrison and Stevens 1976] may be appropriate.

The commentaries on the M-Cornpetition [Armstrong and Lusk 19831 overlook one other salient aspect of short-term forecasting. Suppose the "expert" selects a two-parameter moving average scheme with single differencing, that is ARIMA (0,1,2), whereas "intuitive man" says "use yesterday's value." Their three (or more) step ahead forecasting mean square errors (MSE) are identical, by *definition.* Thus, even when expert and intuitive man choose different models, their forecast MSEs will often converge as the forecasting horizon increases.

Directions for Further Research

The headings on my checklist are not very different from those suggested by Scott Armstrong, but I believe there are some important differences of emphasis. *Model Selection:* When the forecast being generated is sufficiently important to justify modeling effort rather than the automatic use of a standard technique, identification procedures are needed. If, in due course, these can be built into a computer program, so much the better.

However, I believe such methods will come from further research in the practice *and* theory of time-series rather than purely data-analytic studies.

Use of Experts: The purpose of statistical forecasts is to relieve the expert of tedious extrapolation duties, not to compete with him or her. The emphasis on research in this area should be on combining the different information sets available to the statistician and the manager rather than trying to see who does better with the same information.

Combinations of Forecasts: It is still unclear why combinations of forecasts work as well as they do. However, I believe that work on forecasting errors due to model misspecification will guide us to better forecasting than will omnibus methods applied nonselectively.

Assessing Uncertainty in Forecasts: I am not sure that I understand Armstrong's comments on this topic. Prediction intervals are available for most forecasting methods, although these usually depend on assumptions of normality. If the normality clause is under attack, I would agree that further work is needed.

Outlying Observations: Armstrong mentions the need to adjust for outliers, but does not highlight it as a future research area. However, extreme values may seriously disturb estimation procedures, efficient or otherwise (see Martin [1981] and related papers). Outlier adjustment and the development of outlier-resistant estimators are both important, as is the need for forecasting methods which can isolate and adjust to new outliers. Again, Bayesian forecasting [Harrison and Stevens 1976] offers potential solutions here.

Trend Analysis: Since Scott Armstrong's paper emphasizes short-term forecasting, longer term trend analysis may seem out of place in this discussion. However, I think that the lessons we have learned while building short-term models can be applied to longer term forecasting using growth curves [Ord and Young 1984].

Conclusions

Overall, my view of what needs to be done is not so very different from that presented by Scott Armstrong. However, I believe that time-series analysis has contributed fundamentally to the field of forecasting and will continue to do so in the future. Thus, it would appear that our search for solutions will lead us down different pathways which, as the Cheshire Cat might have said, "is just as well if you wand to go to different places."

References

Armstrong, J. Scott (1978), "Forecasting with econometric methods: Folklore versus fact," *Journal of Business*, 51 (4), 549-564 (plus discussion).

Armstrong, J. Scott and Lusk, Edward J. (1983), "The accuracy of alternative extrapolation models: Analysis of a forecasting competition through open peer review," *Journal of Forecasting*, 2 (3), 259-262, with commentary and replies, 263-311.

Box, George E. P. and Jenkins, Gwilym M.(1970), *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Box, George E. P. and Tiao, George C. (1975), "Intervention analysis with application to economic and environmental problems," *Journal of the American Statistical Association,* 70 (1), 70-79.

Harrison, P. Jeffrey and Stevens, Colin F. (1976), "Bayesian forecasting," *Journal of the Royal Statistical Society: Series* B., 38 (3), 205-228 (with discussion).

Makridakis, Spyros, Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1 (2),111-153.

Makridakis, Spyros, and Hibon, Michele (1979), "Accuracy of forecasting: An empirical investigation," *Journal of the Royal Statistical Society: Series A*, 142 (2), 97-145 (with discussion).

Martin, R. Douglas (1981), "Robust methods for time series," *in Applied Time Series,* Vol., II, D. F. Findley, editor, New York: Academic Press, 683-759.

Ord, J. Keith and Young, Margaret R. (1984), "Technological forecasting: Growth curves and discounted least squares," paper to be presented to Fourth International Symposium on Forecasting, London, July 1984.

_____

Reply to the Commentary

My review focused solely on extrapolation methods for short-range forecasting. I agree with Christ that the incorporation of causal factors should improve accuracy for longer-range forecasts. However, more research should be done on that topic.

Ayres referred to the studies in my review as "rather academic." This opinion differs from my own viewpoint, which is that these studies are "valuable" and "interesting." They represent an exciting movement in forecasting away from *description* ("Here is a sophisticated model with some wonderful capabilities.") to *empirical testing* ("Which method is most effective in 'a given situation?"). This is a natural movement in the scientific study of forecasting. I hope it continues. Without it, obviously, there could be no empirical review studies.

The continued growth of empirical testing would help to address the many key issues which are well formulated by Ord. For example, it would be useful to have studies that test Box-Jenkins against alternative procedure in conditions where the B-J procedures are hypothesized to be superior (for example, situations where one-step-ahead forecasts are desired and where ample historical data are available). Finally, I like Ord's call for research studies that combine time series theory with empirical testing.