

Expert Opinions About Extrapolation And The Mystery Of The Overlooked Discontinuities

Fred Collopy The Weatherhead School,
Case Western Reserve University, Cleveland, OH 44106, USA

J. Scott Armstrong The Wharton School,
University of Pennsylvania, Philadelphia, PA 19104, USA

Reprinted with permission *International Journal of Forecasting* 8 (1992) 575-582

Abstract: We report on the opinions of 49 forecasting experts on guidelines for extrapolation methods. They agreed that seasonality, trend, aggregation, and discontinuities were key features to use for selecting extrapolation methods. The strong agreement about the importance of discontinuities was surprising because this topic has been largely ignored in the forecasting literature.

Keywords: Discontinuities, Forecasting guidelines, Selection criteria, Time series features.

1. Introduction

What procedures should be followed in the selection and use of extrapolation methods? Answers to this question should be useful for obtaining accurate forecasts. One way to determine the most appropriate procedures for extrapolation is to ask forecasting experts. This paper reports on the opinions of some experts.

Identifying areas of uncertainty or disagreement among experts, or disagreements between researchers and practitioners, could help to guide further research. Also, the opinions might aid in the development of expert systems for forecasting.

2. Study design

This section describes the questionnaire, our procedures, and the sample of experts that responded.

Our development of the items relied to a great extent upon a review of prior studies, including Mentzer and Cox (1984) and Dalrymple (1987). We also sought advice from forecasting experts and examined the literature in search of guidelines for extrapolation.

The questionnaire asked experts about their roles and experiences, and about what criteria they would use to select forecasting methods. The primary focus of the questionnaire was on features related to the selection of an extrapolation method (10 items) and guidelines for extrapolation (12 items). We sent out the questionnaire in May 1988 with a pre-addressed

stamped return envelope included. One week later we sent a follow-up letter to those who had not responded.

The experts were respondents from a 21% random sample of people who made presentations at the International Symposium on Forecasting in 1987. Of the 100 questionnaires sent out, we received 49 completed questionnaires. (Item non-response was less than 4%.) This number of expert responses is adequate for assessing expert opinions [Armstrong (1985, pp. 85 and 96)]

Most respondents identified themselves as researchers. Eight-three percent said that they did research on forecasting, 56% were forecasting practitioners, and 33% use forecasts in their decision-making. (Respondents were asked to identify all that apply, so the total exceeds 100 %.)

To assess the breadth of the experts' experience, we asked: 'Which of the following methods have you found useful in solving forecasting problems?' The methods, along with the percentage of respondents who found them useful, are listed in Exhibit 1. While the respondents have had experience with a variety of forecasting methods, their primary experience was with extrapolation methods.

Exhibit 1
Forecasting methods found useful by experts

Extrapolation	% Useful	Judgment	% Useful
Simple regression	63	Judgmental	42
Box-Jenkins	54	Scenarios	33
Single exponential smoothing	40	Panel of experts	23
Simple moving average	35	Delphi	17
Holt's exponential smoothing	29	Expert systems	15
Random walk	29	and bootstrapping	
Winter's method	27	Role playing	8
Classical decomposition	25		
Deseasonalization	23	<i>Other</i>	
Bayesian method	19		
Quadratic regression	17	Econometric	56
Census X-11	15	Combining	44
Adaptive response rate exponential smoothing	15		
Gardner's damped trend exponential smoothing	15		
Brown's exponential smoothing	13		
Quadratic exponential smoothing	8		
Lewandowski's FORSYS	6		
Automatic AEP filtering	4		
Parzen's method	4		

Interestingly, the two most frequently mentioned 'useful' methods were simple regression and Box-Jenkins. These have not performed well in a number of published empirical comparisons [Makridakis et al. (1982), Armstrong (1985, pp. 174-178)]. Our experts' general

conclusion about the usefulness of Box-Jenkins differs substantially from that of practitioners; according to Dalrymple (1987), Box-Jenkins was infrequently adopted by practitioners and often discarded after an initial trial. Lowest on our experts' list were three methods that performed well in the M-competition [Makridakis et al. (1982)]: Lewandowski's FORSYS, automatic AEP, and Parzen's method. Perhaps the low ratings occurred because complete and understandable descriptions of these methods were not widely available, and few software packages contained these methods.

When asked 'Relative to other considerations (e.g. cost, ease of interpretation, cost/time, ease of use), how important is the accuracy of the forecasting methods that you use?' 29% of the experts said that accuracy was 'extremely important' and an additional 56% said that it was 'important'. These results are similar to the opinions of practitioners and researchers as reported in Carbone and Armstrong (1982) and with those of practitioners as reported by Mentzer and Cox (1984).

Half the experts said that it was important 'that your forecasting methods provide confidence bounds on the forecasts', while 20% said this was not important. This result contrasts with that from Dalrymple (1987) where only 21% of the firms said that they used confidence intervals 'usually' or 'frequently'.

Most respondents (73%) said that it was 'extremely important' to understand the details and assumption of the forecasting method'. Only 4% (two respondents) thought that this level of understanding was unimportant.

About half of the respondents with opinions said that it was 'moderately important or 'important' that their 'forecasting methods are automatic (computer provided).'

One of the purposes of the study was to identify differences between practitioners and academics' opinions. Operationally, we grouped those who said they were practitioners in one group (n=27), and all others were put in the academics' group (n = 22). Although we had no hypotheses about differences, we assumed that experts with more practical experience would advocate different guidelines. We calculated tests using a critical value for p of 0.05 and making adjustments for multiple comparisons [using the table from Armstrong (1985, p. 467)]. There were no statistically significant differences for the nine features used to select extrapolation methods nor for the 12 guidelines for extrapolation. As a result, we have combined the opinions of all experts in the results below.

3. Results

This section reports on the features used to select extrapolation methods and on the guidelines that experts suggested for extrapolation.

3.1. Features for selecting among extrapolation methods

We examined the features of a time series that should be considered in the selection and use of extrapolation methods. Nine possible features were described (seasonality was defined by two items). The exact wording of these items is provided in Appendix A. Although we expected these features to be important, the experts often had no opinions about them. The percentage of respondents reporting that they ‘neither agree nor disagree’ ranged from 19% to 51% across these items, with an average of 31%. Perhaps the respondents use few features in their selection process or perhaps they do not think about the problem in these terms. Exhibit 2 presents these results by item in the column labeled ‘no opinion.’

Exhibit 2
Features useful in the selection of extrapolation methods

Feature	Percent no opinion	Level of agreement ^a	Percent agreement ^b
Seasonality	32	2.14	98
Recent trend for short forecasts	21	2.09	97
Abrupt changes	19	2.13	92
Aggregation of data	30	2.30	88
Presence of trend	33	2.49	82
Overall trend for long forecasts	23	2.36	81
Overall trend for short forecasts	34	2.64	73
Limits in data	51	2.93	62
Recent trend for long forecasts	33	2.76	60

^a Scale: 1 = strongly agree to 5 = strongly disagree.

^b Omits responses with no opinion.

Exhibit 2 also shows the ‘level of agreement’. The level was calculated by taking the average response (number of responses times the scale for each category divided by the total number of responses). Finally, the last column of Exhibit 2 gives the ‘percent agreement’. This is the number who ‘strongly agreed’ or ‘agreed’ divided by all respondents who had an opinion (i.e. those who answered ‘neither agree nor disagree’ were excluded).

Seasonality and recent trend were regarded as important by nearly all those who had an opinion. These findings were not surprising because existing extrapolation methods are designed to deal with seasonality and trend. The experts indicated that, in addition to the recent trend, the overall trend in a series is an important feature to consider when selecting an extrapolation method. The experts’ opinions about trends can be summarized as follows. The recent trend is most appropriate for short-range forecasting (97% agreed), but fewer experts (60%) thought it was relevant for long-range forecasts. Conversely, the overall trend is relatively important for long-range forecasts (81% agreed) and many (73%) thought it was useful for short-range forecasts.

The feature with the next highest level of agreement was ‘abrupt changes’, with 92% of the experts agreeing that it is important. This is surprising given that time series forecasting

research and practice have largely ignored abrupt changes. To assess this, we examined a convenience sample of 28 books that discuss time series forecasting. This sample was comprised of all relevant books in the second author's (Armstrong's) library. We looked for any use of abrupt changes by examining the books' indices for 'abrupt changes', 'discontinuities', 'erratic fluctuations', 'interruptions', 'irregularities', 'ramps', 'shifts', 'steps', and variations on these terms. Almost all of these books including Armstrong (1985)] completely ignored abrupt changes.¹ Gilchrist (1976) mentioned transients as a feature, but provided no guidelines. Makridakis and Wheelwright (1989) mentioned ramps and provided a guideline on how to handle ramps to the effect that if the monitoring system detects a ramp, then the model should be revised. Our failure to find discussions on the identification and treatment of abrupt changes might be due to our sample or to poorly prepared indices. Nevertheless, this examination of the books shows that little attention has been given to the identification and treatment of abrupt changes. We do not know why such discontinuities have been ignored.

In general, forecasting software has also ignored discontinuities, although there have been exceptions. Bayesian forecasters had developed explicit procedures for monitoring trend and level discontinuities [Harrison and Stevens (1971)], but the approach has not been widely adopted. Coopersmith (1983) and Tsay (1988) concluded that level changes are often ignored and recommended procedures for dealing with them. Recent developments in some Box-Jenkins forecasting packages have included facilities for dealing with abrupt changes.

The level of aggregation of the data was expected to be related to the relative accuracy of alternative extrapolation methods by 88% of the experts. We speculate that the level of aggregation may be important because different causal factors might affect different components. Highly aggregated data are more likely to be subject to different causal factors than are less aggregated data. On the other hand, the reliability of data often improves when one uses larger aggregates.

The respondents were ambivalent about whether limits in the data (e.g. market share data) should affect the selection of a method. Advocates of diffusion models, for example, use asymptotic limits when they forecast the growth of sales for new products. This ambivalence is probably warranted, as we are not aware of research demonstrating the importance of limits for extrapolation models.

The features identified by this study were useful in the development of a rule-based forecasting system [Collopy and Armstrong (1992)]. In particular, rule-based forecasting benefited from the emphasis that the experts placed on abrupt changes. To deal with abrupt changes, we identified time series that had level discontinuities and trend discontinuities. Our identification procedures were reliable; the intercoder reliability between two raters was 91% for level discontinuities and 81% for trend discontinuities. We developed rules for series with abrupt changes. The presence of abrupt changes played a role in 11 of 99 rules in our rule base. These rules contributed to substantially improved accuracy when applied to new sets of time series.

¹ The term 'irregularities' was used as a synonym for 'random error' by some authors.

3.2 Guidelines for extrapolation

What advice do experts offer about making extrapolation forecasts? Exhibit 3 indicates the level of agreement among the experts for 12 guidelines. The guidelines are arranged in order of agreement (last column). The exact wording of these items is provided in Appendix B.

The first guideline in Exhibit 3 is especially relevant to the users of extrapolation methods. It shows that 83% of the experts with an opinion believe that combining will produce more accurate forecasts. The experts believed that, in general, combined forecasts are more accurate than those based on a single method: 73% of the respondents agreed and only 15% disagreed. These beliefs are consistent with empirical research [reviewed in Clemen (1989)]. The experts were undecided about whether combining was more appropriate for long- or short-range forecasts. Empirical research, though limited, suggests the latter [Lawrence, Edmundson and O'Connor (1986)].

Exhibit 3
Guidelines useful for extrapolation

Guidelines to improve accuracy	Percent no opinion	Level of agreement ^a	Percent agreement ^b
Combine forecasts	13	2.13	83
Use adaptive forecasts for short horizons	30	2.53	75
Combine judgmental forecasts with extrapolation	43	2.64	73
Adjust current status by judgment	43	2.68	73
Use judgment, not computer, to adjust outliers ^c	43	2.70	69
Use simpler models for unstable series	44	2.79	69
Adjust trend by judgment	50	2.83	65
Use all available data points	29	2.84	59
Use simpler methods for stable series	42	2.94	59
Use simpler methods if uncertainty is high	44	2.90	58
Combining more important for long range	47	2.94	56
Do not use adaptive forecasts for long range ^c	34	2.98	53

^a Scale: 1 = strongly agree to 5 = strongly disagree.

^b Omits responses with no opinion

^c The wording of these questions was reversed for presentation

Seventy-three percent of the experts believe that improved accuracy can be gained by combining judgment with extrapolation. Surprisingly, it is difficult to find studies that have examined the benefits of combining judgment forecasts with extrapolation forecasts. The guideline is, however, consistent with the limited prior evidence based on a reanalysis of two studies by Armstrong (1985, pp. 290-292) and a study by Blattburg and Hoch (1990).

While the guideline to use adaptive forecasts for short horizons had 75% agreement, it is not supported by empirical evidence. See Armstrong (1985, p. 171) for a review of this evidence. Although this research suggests that adaptive forecasts did not improve accuracy, we suspect that adaptive forecasts may be relevant for some situations, such as for series with discontinuities.

Judgmental adjustments of the level and trend were expected to improve accuracy. The empirical research supports adjustments of levels, but trend adjustments are controversial [Armstrong (1985, pp. 235-238)]. To some degree, the experts agreed that judgmental adjustments were more appropriate for the current level than for the trend.

The respondents favored simple methods for stable and unstable situations, with a slightly stronger preference for their use in unstable situations. Schnaars' (1986) results implied that simple models are most appropriate for unstable situations.

Most respondents believed that more data points would increase accuracy, but only three respondents (6%) strongly agreed. Some experts believed that the amount of data was not important, and 29% said that they 'neither agree nor disagree' that more data improves accuracy. We found the expert opinions on guidelines in the upper half of Exhibit 3 to be valuable in developing a rule-based forecasting system [Collopy and Armstrong (1992)]. For example, as the amount of instability in a time series increases, we place an increasing emphasis on the simplest method, the random walk.

The experts were ambivalent about many of the guidelines. Across the 12 items, respondents selected 'neither agree nor disagree' 38% of the time. Two respondents expressed annoyance at this line of questioning, perhaps feeling that the responses depend upon the situation.

The guidelines at the bottom of Exhibit 3 are those on which there was much disagreement among the experts as to the best procedure. For example, we originally believed that truncation of data was clearly a useful strategy, but this belief received little support from the experts. Further research might resolve these conflicting opinions.

The direct assessment of expert opinions provides general guidelines, but it does not provide sufficient detail for the development of rules for forecasting. This detail is more effectively obtained from protocol analyses [Collopy and Armstrong (1989)].

4. Conclusions

This paper summarizes some opinions of forecasting experts. We identified features of time series to be considered in the selection of an accurate extrapolation method. Not surprisingly, the experts agreed that seasonality and recent trend were key features. The experts also placed a heavy emphasis on the importance of abrupt changes in the historical data patterns. This stands in stark contrast to forecasting methods and forecasting research which have long ignored abrupt changes. We have no explanation for this mystery of the overlooked discontinuities. Fortunately, software developers are responding to this problem.

Although they agreed on the value of combining forecasts, the experts had diverse opinions about what additional guidelines should be used for extrapolation forecasts. The study identified many areas where the experts were undecided. These areas of disagreement and uncertainty might be worthy of further study.

Acknowledgments

The US Navy Personnel R&D Center in San Diego provided partial support for this research. We thank James E. Cox, Timothy A. Davidson, Douglas J. Dalrymple, Robert Fildes, Leonard J. Tashman, J. Thomas Yokum, and two anonymous referees for their comments.

Appendix A: Items used to determine features useful in the selection of extrapolation methods

Seasonality: The presence of seasonality in the data series is significantly related to differences in the accuracy of alternative extrapolation methods. The magnitude of seasonality in the data series is significantly related to differences in the accuracy of alternative extrapolation methods.

Recent trend for short forecasts: The magnitude and type of the recent trend in the data series is, for short-range forecasts, significantly related to differences in the accuracy of alternative extrapolation methods.

Abrupt changes: Abrupt pattern changes in the data series are significantly related to differences in the accuracy of alternative extrapolation methods.

Aggregation of data: The level of aggregation of the data (e.g. macro/micro or product class/product form) is significantly related to the relative accuracy of alternative extrapolation methods.

Presence of trend: The presence of trend in the data series is significantly related to differences in the accuracy of alternative extrapolation methods.

Overall trend for long forecasts: The magnitude and type of the overall trend in the data series is, for long-range forecasts, significantly related to differences in the accuracy of alternative extrapolation methods.

Overall trend for short forecasts: The magnitude and type of the overall trend in the data series is, for short-range forecasts, significantly related to differences in the accuracy of alternative extrapolation methods.

Limits in data: The existence of limits such as percentages in a data series (e.g. forecasting market share) is significantly related to the accuracy of alternative extrapolation methods.

Recent trend for long forecasts: The magnitude and type of the recent trend in the data series is, for long-range forecasts, significantly related to differences in the accuracy of alternative extrapolation methods.

Appendix B: Items used to determine guidelines useful for extrapolation

Combine forecasts: Using combinations of forecasts from multiple extrapolation methods will provide greater accuracy than relying on a single method.

Use adaptive forecasts for short horizons: Models that readily adapt to change are relatively more accurate in forecasting short horizons than those that do not.

Combine judgmental forecasts with extrapolation: Combinations that include a forecast done by judgmental methods will be more accurate than those that do not.

Adjust current status by judgment: Forecasts for which the current status has been adjusted by judgment will be more accurate than those for which it has not.

Use judgment, not computer, to adjust outliers: Series with outliers are better forecast by methods that incorporate facilities for automatically dealing with them than by those that rely on the forecaster making these adjustments.

Use simpler models for unstable series: Simple models are more accurate in forecasting unstable series than more complex models.

Adjust trend by judgment: Forecasts for which the trend has been adjusted by judgment will be more accurate than those for which it has not.

Use all available data points: The greater the number of data points used in selecting and fitting a model the more accurate will be the forecasts.

Use simpler models for stable series: Simple models are more accurate in forecasting stable series than complex models.

Use simpler models if uncertainty is high: Simpler models will more accurately forecast series with a high level of randomness than more statistically sophisticated models.

Combining more important for long range: Improvements in accuracy as a result of combining forecasts will be greater for the long term than for the short term.

Do not use adaptive forecasts for long range: Models that readily adapt to change are relatively more accurate in forecasting long horizons than those that do not.

References

Armstrong, J.S., 1985, *Long-Range Forecasting* (John Wiley, New York).

- Blattburg, R.C. and S. J. Hoch, 1990, "Database models and managerial intuition: 50% model + 50% manager," *Management Science*, 36, 887-899.
- Carbone, R. and J.S. Armstrong, 1982, "Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners," *Journal of Forecasting*, 1, 215-217.
- Clemen, R.T., 1989, "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559-583.
- Collopy, F. and J.S. Armstrong, 1989, "Toward computer-aided forecasting systems," in: G.R. Widmeyer, ea., *DSS 89 Transactions* (TIMS College on Information Systems, Providence, RI), 103-119.
- Collopy, F. and J.S. Armstrong, 1992, "Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations," *Management Science*, 38.
- Coopersmith, L.W., 1983, "Forecasting time series which are inherently discontinuous," *Journal of Forecasting*, 2, 225-235.
- Dalrymple, D.J., 1987 "Sales forecasting practices," *International Journal of Forecasting*, 3, 379-391.
- Gilchrist, W., 1976, *Statistical Forecasting* (John Wiley, London).
- Harrison, P.J. and C.F. Stevens, 1971, "A Bayesian approach to short-term forecasting," *Operational Research Quarterly*, 22, 341-362.
- Lawrence, M.J., R.H. Edmundson and M.J. O'Connor, 1986, "The accuracy of combining judgmental and statistical forecasts," *Management Science*, 32, 1521-1532.
- Makridakis, S. and S. C. Wheelwright, 1989, *Forecasting Methods for Management* (John Wiley, New York).
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, 1982, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.
- Mentzer, J.T. and J.E. Cox, 1984, "Familiarity, application and performance of sales forecasting techniques," *Journal of Forecasting*, 3, 27-36.
- Schnaars, S.P., 1986, "Situational factors affecting forecast accuracy," *Journal of Marketing Research*, 21, 290-297.
- Tsay, R. 1988, "Outliers, level shifts, and variance changes in time series," *Journal of Forecasting*, 7, 1-20.