Reprinted with permission from *Journal of Forecasting*, Vol. 1, 215-1217 (1982).

Evaluation of Extrapolative Forecasting Methods: Results of a Survey of Academicians and Practitioners

Robert Carbone
Faculté des sciences de l'administration, Université Laval

J . Scott Armstrong
The Wharton School, University of Pennsylvania

There exists a large number of quantitative extrapolative forecasting methods which may be applied in research work or implemented in an organizational setting. For instance, the lead article of this issue of the *Journal of Forecasting* compares the ability to forecast the future of over twenty univariate forecasting methods. Forecasting researchers in various academic disciplines as well as practitioners in private or public organizations are commonly faced with the problem of evaluating forecasting methods and ultimately selecting one. Thereafter, most become advocates of the method they have selected. On what basis are choices made? More specifically, what are the criteria used or the dimensions judged important? If a survey was taken among academicians and practitioners, would the same criteria arise? Would they be weighted equally? Before you continue reading this note, write on a piece of paper your criteria in order of importance and answer the last two questions. This will enable you to see whether or not you share the same values as your colleagues and test the accuracy of your perception.

The general meeting of the First International Symposium on Forecasting held last May in Quebec City offered a unique occasion to survey forecasting practitioners and academicians on these issues. The conference attracted over 500 participants from more than 20 countries. Those who attended the general meeting were asked to list on a sheet of paper the criteria they felt should be used for evaluating extrapolative forecasting methods. They were also asked to identify themselves as either a practitioner or an academician. Specific responses were requested. For example, if accuracy was deemed an important criterion, then what measure(s) should be used for evaluating accuracy.

Two hundred and six attendees submitted a response. However, only 145 replies were retained; 75 practitioners and 70 academicians. The remaining ones were discarded either because (i) the respondent did not identify himself (90 per cent of the replies discarded) or (ii) the reply addressed other issues (for example, selection of independent variables in causal modeling).

Table I lists the criteria and gives a tabulation of the number of times each criterion was reported. Under the accuracy criterion, the number of times a specific measure was stated is also indicated.

Table 1. Evaluative criteria and their relative importance as determined by forecasting practitioners and academicians

Criteria	Academicians	Practitioners
	70	75
Accuracy		
R^2		2
Mean square error (MSE)	30	20
Geometric MSE	1	
Minimum variance	2	4
Theil's U test	3	1
Mean percentage error (MPE)	5	5
Mean absolute error (MAE)	12	14
Mean absolute percentage error (MAPE)	15	7
Minimax absolute error (MMAE)	2	
Random forecast errors	1	2
No specific measure	8	14
Ease of interpretation	26	29
Cost/time	24	25
Ease of use/implementation	26	18
Adaptive to new conditions	10	13
Universality	3	10
Capture turning points	5	6
Robustness	10	3
Incorporates judgmental input	4	2

It is not too surprising to find from this table that accuracy, ease of interpretation, cost/time, ease of use/implementation and adaptive to new conditions were most frequently reported. The most important among them is without any doubt accuracy, with only 14 per cent of the respondents (7 academicians and 12 practitioners) excluding it as a criterion. What may be more interesting to note when looking at the table is the similarity in the responses given by academicians and practitioners. For example of those who indicated an accuracy measure, the square error norm dominates in the two groups in about the same proportion. It is also interesting to note that the mean absolute percentage error did not reveal itself the favorite measure among practitioners. In fact, it ranks after mean absolute error and twice as many academicians listed it.

If differences are to be found, they reside in the least cited criteria. For example, universality (applicable to different situations, types of series or environments) is more highly weighted by practitioners. In contrast, robustness is viewed as more important to academicians. Of importance also are those criteria which have only received lip service by both groups. Random forecast errors is surprisingly one of them. R^2 is another.

Even if the general conclusion of this survey is that the same criteria are used by both groups, a certain lack of agreement still exists within each group. This survey in fact raises more questions than it answers

1. Would a closed-ended questionnaire produce similar results to the open question asked?
2. Are the expressed preferences in agreement with actual preferences?

3. What criteria are actually presented in reports by academicians and practitioners?
4. Are the expressed preferences reasonable ? In other words, will the preferences provide the best guide to decision making? What criteria would be the consensus choice among forecasters and decision makers?
5. Does the selection of the criteria depend upon the situation? If so, in what way?
6. Does the selection of the criteria depend upon the role of the evaluator, e.g. forecaster vs. decision-maker?
7. What criteria are proposed in texts on forecasting methods?

Hopefully, study of these questions will lead to a set of accepted criteria.