

Beyond Accuracy: Comparison of Criteria Used to Select Forecasting Methods

J. Thomas Yokum^a

J. Scott Armstrong^b

Abstract

Two studies were conducted to examine expert opinions of criteria used to select forecasting techniques. In Study One, while "accuracy" was a dominant criterion, the ratings of five of thirteen criteria varied by the role of the forecaster. Researchers rated accuracy relatively higher than did practitioners, educators or decision-makers. Decision makers rated implementation-related criteria, such as "ease" criteria, relatively higher than the other groups. In Study Two, forecasting experts significantly varied their ratings on six of seven criteria according to situations. Other criteria were often as important or more important than accuracy, especially when the situation involved making many forecasts. In general, there was much agreement across roles and across situations that accuracy was the most important criterion, but other criteria were rated as being almost as important. In particular, factors related to implementation, such as ease of interpretation and ease of use, were highly rated.

Keywords: Expert opinion; Forecaster's role; Forecast situation; Implementation

1. Introduction

Research in forecasting has commonly assumed that accuracy is the primary criterion in selecting among forecasting techniques. In fact, it has been used as the *sole* criterion in many studies. In the sixteen 1992 *International Journal of Forecasting* papers that compared the results of different techniques and series, only one used criteria other than accuracy.¹ Is this single-minded focus on accuracy reasonable? In this paper, we examine the importance of various criteria. We also examine whether the various roles of forecasters would lead them to prefer different criteria and whether the nature of the forecasting situation would affect the selection criteria. We address these issues by first examining prior research, then by obtaining opinions of forecasting experts in two complementary studies.

2. Prior Studies

In Collopy and Armstrong's (1992) study on the opinions of 49 forecasting experts, 29% of their respondents rated accuracy as "extremely important" relative to other criteria such as ease of interpretation, cost/time, and ease of use. An additional 56% rated accuracy as "important." In all then, 85% of the experts rated accuracy as an important criterion.

Four studies compared accuracy with other criteria. Each concluded that accuracy was the most important criterion for evaluating forecasting techniques, but they produced different rankings for other criteria (see Table 1). Business decision makers, the only respondents in both the Mentzer and Cox (1984) and the Mahmoud et al. (1986) studies, ranked "ease of use" second. In Carbone and Armstrong (1982), academics and practitioners ranked "ease of

^a Angelo State University, San Angelo, TX 76909, USA

^b Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA

¹ Scott (1992) compared four X11 techniques on their ease of use.

interpretation" second, while practitioners ranked "cost/time" third and academics ranked "ease of use" second. Witt and Witt (1992) surveyed tourism academics, government officials, and practitioners on four criteria evaluated over short, medium, and long term horizons. Although accuracy was generally rated first overall, government officials rated "speed" first for the short-term; practitioners ranked "ease of use" first for long-term forecasts. Rankings for all studies are given in Table 1.

The limited evidence indicates that the selection of different criteria may depend upon the rates and the conditions under which they were selected. Carbone and Armstrong (1982) found modest differences between academics and practitioners, while Witt and Witt (1992) found some differences in their short- and long-term horizon criteria rankings.

Table 1. Rankings of criteria from previous studies (number of respondents)

	Carbone and Armstrong, 1982 (145)	Mentzer and Cox, 1984 (160)	Mahmoud et al., 1986 (67)	Witt and Witt, 1992 (29)
Accuracy	1	1	1	1
Ease of interpretation	2	—	—	—
Ease of use	4	2	2	2
Ease of using data	—	—	3	—
Credibility	—	3	—	—
Speed	—	—	—	3
Cost savings	3	4	—	4
Ease of implementation	4	—	—	—
Time horizon	—	—	4	—
Adaptive to conditions	5	—	—	—

3. Study One: Expert Opinions by Role

In Study One, we assessed whether criteria differ according to the respondent's primary occupational role. The criteria we used were selected from a combination of those from previous studies and from discussions with a number of leading forecasters. Descriptions for the criteria, following the format of the four previous studies, are listed in Table 2. Although we believed that some of the phrases might be considered to be a bit ambiguous, respondents did not voice concerns about the definitions in either our pretest or our study.

Table 2. Importance of criteria in selecting a forecasting technique (scale – 1 “unimportant” to 7 important”)

Question	Mean agreement rating				
	Avg.	Decision maker (DM)	Practitioner (PR)	Educator (ED)	Researcher (RS)
Accuracy	6.20	6.20	6.10	6.09	6.39* ^{DM, PR, ED}
Timeliness in providing forecasts	5.89	5.97	5.92	5.82	5.87
Cost savings resulting from improved decisions	5.75	5.97	5.62	5.66	5.89
Ease of interpretation	5.69	5.82	5.67	5.89	5.54
Flexibility	5.58	5.85* ^{PR, ED, RS}	5.63	5.35	5.54
Ease in using available data	5.54	5.79	5.44	5.52	5.59
Ease of use	5.54	5.84* ^{PR, RS}	5.39	5.77* ^{PR, RS}	5.47
Ease of implementation	5.41	5.80* ^{PR, ED, RS}	5.36	5.55	5.24
Incorporating judgmental input	5.11	5.15	5.19	5.12	4.98
Reliability of confidence int.	4.90	5.05	4.81	4.70	5.09
Development cost (computer, human resources)	4.86	5.10	4.83	5.02	4.70
Maintenance cost (data storage, modifications)	4.73	4.72	4.73	4.75	4.71
Theoretical relevance	4.40	3.72	4.43* ^{DM}	4.20* ^{DM}	4.81* ^{DM}

* denotes significantly higher ratings ($p < 0.05$) for column group versus group/s listed in superscript.

We received replies from 322 experts from a total of 738 questionnaires sent to International Institute of Forecasting (IIF) members and nonmembers.² The respondents were divided into four groups based on self-ratings of their primary forecasting role. They were asked to select which of the role descriptions best fit them (see Appendix). There were 94 researchers, 55 educators, 133 practitioners (i.e. forecast preparers), and 40 decision makers (i.e. forecast users). Each group is well above the range of the number of experts (5 to 20) thought to be needed for the analysis of expert opinion (Hogarth, 1978).

One concern was that the level of expertise of our respondents might lead to different judgments. To assess this possibility, we compared the responses of 269 IIF members with 53 nonmembers, who may have less expertise. Using ANOVA, we found that membership status did not make a difference in importance rankings for any of the 13 criteria. Thus, responses from all experts were combined.

"Accuracy" was rated most important by all four role groups. It had an average rating of 6.20 on a seven-point scale. As noted, accuracy had been the most important criterion in the previous studies. To our surprise, it remained so despite the addition of the criterion "cost savings resulting from improved decisions." The second most important criterion (average rating 5.89) over all groups was "timeliness in providing forecasts." This criterion had not been mentioned in previous studies.

One might expect that decision makers would be looking at the ultimate benefits of forecasting, such as "cost savings resulting from improved decisions." This was only rated third overall (average rating 5.75).

Following "cost savings from improved decisions" were five criteria that emphasized "ease" of some particular forecasting facet. The "ease of interpretation" element was rated highest, followed by "flexibility," "ease in using available data," "ease of use," and "ease of implementation." The five ease components were similarly rated with average ratings ranging from 5.41 to 5.69. The importance of ease is consistent with the finding in Sanders and Manrodt's (1994) survey on sales forecasting in large U.S. firms. In their study, "ease of use" was rated as the most important advance their respondents would like to see with respect to forecasting.

Note that across all respondents, the costs of developing and maintaining the forecasting system were ranked as relatively less important criteria.

The lowest rated criterion was "theoretical relevance" (average rating 4.40). We suspect that this may be due to the assumption that one has the choice of many models, each of which may be perceived to be theoretically relevant in a particular situation.

In general, the experts rated all criteria as being relatively important. In fact, there was no criterion average below 4, the midpoint of the importance scale used in the study.

Duncan tests in the ANOVA analysis indicated that statistically significant group differences in Table 2 occurred for five of the thirteen criteria. The main differences were between decision makers and the other three forecasting groups. Decision makers ranked "flexibility" and "ease of implementation" significantly higher than the other three groups, and decision makers ranked "ease of use" higher than did practitioners and researchers. Additionally, they were significantly lower in their rankings of "theoretical relevance." Decision makers appear to be especially concerned about what technique will work well over various work-related situations. Researchers, meanwhile, placed emphasis on "accuracy," rating it significantly higher than did the other three groups.

Differences between researchers and decision makers were confirmed by the relative order of the criteria. Researchers rated "theoretical relevance" highest of the groups, while rating "development costs," "maintenance costs," "incorporating judgmental input," "ease of use," and "ease of interpretation" lower than any of the other groups. These ratings suggested that researchers desired more quantitative-based comparison of forecasting techniques, versus the more implementation-related "ease" comparisons for decision makers.

Although there were differences among forecasters with different roles, the basic story was one of agreement. Three of the top four rated criteria ("timeliness in providing forecasts," "cost savings," and "ease of interpretation") did not show differences among the four groups. Other similar group ratings included "ease in using available data," "incorporating judgmental input," "reliability of confidence interval," "development cost," and

² Expert surveys differ from attitude and behavior surveys in that the experts are asked their opinions about phenomena outside themselves. There is no attempt to generalize about the respondents themselves.

"maintenance cost." Correlations among the average group ratings of criteria in Table 2 had an average of 0.90. The range between the groups was from 0.81 for researcher's and decision maker's ratings to 0.95 for the ratings between educators and decision makers.

In conclusion, Study One shows "accuracy" to be the dominant criterion, as it was rated highest by all groups, and significantly higher by researchers. The role of a forecaster had some impact on the selection of criteria. The most important findings, however, were (1) some criteria were rated as being almost as important as accuracy, and (2) although there was some role variation in five criteria ratings, there was much agreement across roles as to the importance of various criteria.

4. Study Two: Expert opinions by situation

Forecast method selection criteria might also vary by the nature of the forecasting situation. For example, Witt and Witt (1992) found that "speed" was most important for short-range forecasts, while "accuracy" was most important for medium- and long-term forecasts. Study Two looked at three factors-the amount of data available (long or short time-series), the number of forecasts required (many or few forecasts), and the method used (extrapolation or econometric models).

Twenty-seven participants at the 1993 International Symposium of Forecasting completed self-administered questionnaires.³ This provided 189 evaluations across seven criteria: "accuracy," "savings from improved decisions," "flexibility," "ease of using available data," "ease of use," "ease of interpretation," and "quickly provides forecasts."⁴ The seven criteria were cross-classified by eight nested (2 x 2 x 2) situations.

Table 3. Criteria ratings – overall averages and averages across situations

Criteria	Avg.	Many forecasts				Few forecasts			
		Long series		Short series		Long series		Short series	
		Extp.	Econ.	Extp.	Econ.	Extp.	Econ.	Extp.	Econ.
Accuracy	5.88	5.74	5.68	5.91	5.59	6.04	6.00	6.25	5.81
Cost savings from improved decisions	5.81	5.54	5.95	5.91	5.87	5.58	5.96	5.92	5.83
Ease of interpretation	5.32	5.00	5.68	4.86	5.41	5.22	5.82	5.08	5.50
Flexibility	4.99	4.91	5.00	4.96	4.86	4.78	5.08	4.79	4.86
Ease of using available data	4.89	5.48	4.50	5.52	4.86	4.83	4.61	5.00	4.41
Ease of use	4.43	4.65	4.50	4.69	4.14	4.26	4.30	4.42	4.09
Quickly provided forecasts	4.12	4.35	4.14	4.87	4.27	3.91	3.78	4.25	3.82

Our primary concern was whether the experts would rate the criteria differently across different situations. The results showed that the situation did make a difference. Respondents varied 61% of the criteria ratings across the eight situations. The percentage of each of the criteria ratings that varied ranged from a low of 28% for "accuracy" to a high of 70% for three criteria: "flexibility," "ease of using available data," and "quickly provides forecasts." We also examined the agreement across the situation by calculating the standard deviation of each criterion. A standard deviation of zero would indicate that respondents gave the same ratings irrespective of the situation. "Quickly provides forecasts" was the criterion that varied the most across situations with a standard deviation of 1.4. The standard deviations for the "ease" factors of interpretation, using available data, use, and flexibility were approximately 1.3. "Accuracy" varied least with a standard deviation of about 1.

The mean ratings of the criteria across situations, listed from highest to lowest, in the first column of Table 3, generally mirror Study One's results. "Accuracy" was rated highest in importance, but again the differences were small. In Study Two, the experts often rated other criteria higher than accuracy for certain situations. On average across the eight situations, about 41 % of the time the respondents rated one of the six other criteria equal to or greater than accuracy. These other criteria were significantly more important when making many forecasts versus making few forecasts (46.4% vs. 35.2%, $t = 2.23$, $p < 0.025$). Non-accuracy criteria were higher in importance, but

³ Questionnaires were handed out to approximately 60 conference participants at a session break.

⁴ These criteria were the same as those in Study One except that "timeliness in providing forecasts" was replaced by the phrase, "quickly provides forecasts." We thought that the original term was vague, but the variation in response for this item was substantial. In short, our attempt at improving upon the previous wording did not seem successful.

not significantly higher, when using econometric versus extrapolation techniques (42.8% vs. 38.8%, $t = .79$, $p < 0.50$).

The relative order of the importance ratings of the criteria in the two studies were generally similar. "Accuracy" was followed by "cost savings," "ease of interpretation," "flexibility," "ease of using available data," and "ease of use." The mean ratings for "quickly provides forecasts," fell from second in importance in the first study to last in the second study. It was thought to have a similar meaning to "timeliness in providing forecasts" from the first part of the study.

The 56 mean ratings of situation by criteria are also listed in Table 3. The patterns of the means show higher importance ratings for "accuracy" when making few forecasts. Additionally, "accuracy," "ease of use," "ease of using available data," and "quickly provides forecasts" were more important when using extrapolation techniques. "Ease of interpretation" was more important for econometric techniques.

5. Conclusions

Previous studies on selection criteria and the two studies that we report here, support the commonly held assumption that accuracy is the most important criterion. This opinion was expressed by experts, regardless of their role and regardless of the forecasting situation. That said, the experts viewed a number of other criteria as roughly comparable in importance. This was especially true for implementation-related factors, including the five ease criteria involving interpretation, use, flexibility, employing available data, and implementation. Thus, from a practical implication perspective, multiple criteria are desirable in the selection and evaluation of forecasting techniques. This would especially be relevant to developers of new forecast techniques. To encourage diffusion, new techniques should be evaluated, not only in terms of comparative accuracy, but also in terms of the "ease of use," "ease of interpretation," and "flexibility."

"Cost savings" varied in rank depending upon its framing from a top criterion if related to savings from improved decisions to a lower criterion if linked to savings from technique development and maintenance.

Our Study One showed that the expert's ratings of five of the thirteen criteria varied according to the forecasting role. Decision makers rated application and implementation criteria such as "flexibility," "ease of implementation," and "ease of use" more highly than did the other groups. Researchers rated "accuracy" significantly higher than did the other three groups. It should be noted that although importance varied by groups, the experts generally agreed on the order of the relative importance of the criteria. The experts appeared to seek savings from good forecasting, not from cutting development and maintenance costs. Theoretical relevance of the technique was rated last.

In Study Two, although "accuracy" did not vary much, six other criteria showed considerable variation across situations. We believe that criteria may be easier to think about in the context of the forecasting situation. There were noticeable differences in means by situation. Especially notable was the result that many of the other criteria were rated equal or greater than the importance of accuracy for some situations. Although it may not be appropriate in all situations, we suggest that researchers go beyond accuracy when comparing forecasting methods. For example, while accuracy is comparable across all techniques, different situations, circumstances, and usages would necessitate different criteria. "Ease of interpretation" would seem to be appropriate for econometric models where estimation of marketing mix elasticities might be more important than accuracy considerations. When comparing techniques, a variety of criteria should be used.

Appendix

A.1. Information about the experts

The initial study of expert opinion was mailed to 521 members of the International Institute of Forecasters and contained over 12 areas of interest to forecasters. Two hundred and sixty-nine responses were returned. An additional 229 surveys were mailed to non-IIF members from a list of former participants in forecasting conferences. Fifty-three responses were returned making a total of 322. The respondents were from 30 countries, with the majority from the U.S., Canada, and Europe. This large number of respondents helped to ensure that we had sufficient numbers of experts when broken down by role and by situation.

A.2. Forecasting role descriptions

Researchers: Individuals who primarily pursue the development and testing of forecasting theory. They disseminate knowledge through professional speeches and research articles.

Educators: Those who formally educate others about forecasting thought and theory. Dissemination of knowledge involves direct contact, usually within the environment of a university, but also through private firms, professional organizations, various seminars, professional meetings, and textbooks.

Practitioners: Those who help to prepare and present forecasts within a profit, nonprofit or public organization. Their knowledge about forecasting is based on their experience, procedures developed by the organization, and the efforts of educators and researchers.

Decision makers: Individuals within a firm who use forecasts to plan, make decisions, implement strategies, and control performance measures. They depend upon the forecasting practitioner.

Acknowledgements

The authors thank Fred Collopy and Essam Mahmoud for their helpful comments.

References

- 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.
- Collopy F. and J.S. Armstrong, 1992, "Expert opinions about extrapolation and the mystery of the overlooked discontinuities," *International Journal of Forecasting* 8, 575-582.
- Hogarth, R.M., 1978, "A note on aggregating opinions," *Organizational Behavior and Human Performance* 21, 40-46.
- Mahmoud E., G. Rice and N. Malhotra, 1986, "Emerging issues in sales forecasting and decision support systems," *Journal of Academy of Marketing Science* 16, 47-61.
- Mentzer and J. Cox, 1984, "Familiarity, application, and performance of sales forecasting techniques," *Journal of Forecasting* 3, 27-36.
- Sanders, N. R. and K. S. Manrodt, 1994, "Forecasting practices in U.S. corporations," *Interfaces* 24, 92-100.
- Scott, S., 1992, "An extended review of the X11ARIMA seasonal adjustment package," *International Journal of Forecasting* 8, 627-633.
- Witt, S. and C. Witt, 1992, *Modeling and Forecasting Demand in Tourism*. Academic Press, London.