2-1986

Research on Forecasting: A Quarter-Century Review, 1960-1984

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Recommended Citation

Armstrong, J. S. (1986). Research on Forecasting: A Quarter-Century Review, 1960-1984. *Interfaces, 16* (1), 89-109. http://dx.doi.org/10.1287/inte.16.1.89

Research on Forecasting: A Quarter-Century Review, 1960-1984

Abstract
Before 1960, little empirical research was done on forecasting methods. Since then, the literature has grown rapidly, especially in the area of judgmental forecasting. This research supports and adds to the forecasting guidelines proposed before 1960, such as the value of combining forecasts. New findings have led to significant gains in our ability to forecast and to help people to use forecasts. What have we reamed about forecasting over the past quarter century? Does recent research provide guidance for making more accurate forecasts, obtaining better assessments of uncertainty, or gaining acceptance of our forecasts? I will first describe forecasting principles that were believed to be the most advanced in 1960. Following that, I will examine the evidence produced since 1960.

Keywords
Forecasting

Disciplines
Business | Marketing

(Reprinted with permission from *Interfaces* (1986)16:1 (January-February), 89-103)

Research on Forecasting: A Quarter-Century Review, 1960-1984

J. Scott Armstrong
The Wharton School of the University of Pennsylvania

Before 1960, little empirical research was done on forecasting methods. Since then, the literature has grown rapidly, especially in the area of judgmental forecasting. This research supports and adds to the forecasting guidelines proposed before 1960, such as the value of combining forecasts. New findings have led to significant gains in our ability to forecast and to help people to use forecasts.

What have we reamed about forecasting over the past quarter century? Does recent research provide guidance for making more accurate forecasts, obtaining better assessments of uncertainty, or gaining acceptance of our forecasts?

I will first describe forecasting principles that were believed to be the most advanced in 1960. Following that, I will examine the evidence produced since 1960.

Forecasting Principles in 1960

Forecasts serve many needs in organizations and are employed for both short-range and long-range planning. They help in making decisions on production, personnel, finance, and marketing. Figure 1 shows the types of forecasts that are needed in addressing these decisions. For example, forecasts of competitive actions can help to assess a proposed strategy.

The same needs for forecasts existed in 1960. I drew upon the pre-1960 literature and upon my experience (I was just starting my life as a forecaster in 1960) to de scribe what methods would have been used in 1960. From this, I developed 1960-style guidelines, six for making forecasts, one for estimating uncertainty in the forecasts, and one for gaining acceptance of the forecasts:

(1) Decomposition should be used whenever feasible. Complex problems should be broken into a series of sub-problems, each of which is to be solved, and the results then synthesized. Of particular importance are decompositions to separate the forecasting of company sales into industry sales and market share, estimate current status separately from forecasting change, and provide forecasts separately for each consumer market. Decomposition can be used with judgment, econometric or extrapolation methods. Its use in extrapolation (whereby average, trend, and seasonal components are examined) is especially popular.

(2) Extrapolation should be used as one of the forecasting methods whenever the data allow. Exponential smoothing of deseasonalized data using a trend estimate will produce adequate results.

(3) Obtain opinions from experts in the topic area, and do this in highly structured ways. Use these opinions to estimate current status, make forecasts of change, forecast in situations where data on the variable of interest (for example, the sales variable) are absent or of poor quality, and forecast the effects of actions by the firm when considering the reactions of major competitors, but do not use expert opinion to adjust the forecasts produced by objective methods.

(4) Intentions surveys (for example, consumers' intentions) should be used to forecast short-term behavior for important events.

(5) Causal-objective methods, such as econometric models, are preferred whenever sufficient data exist. They should be developed from a priori theory, and estimates should be obtained from a variety of sources. Highly reliable data are desirable.

(6) Combine forecasts from at least two different approaches. Each forecast should be weighted according to the confidence one has in it.

(7) Uncertainty estimates should accompany the forecasts. The accuracy of a given method in similar situations (that is, the method's track record) should be used as the primary means for assessing uncertainty. Independent judgmental assessments of accuracy should also be used.

(8) Before the forecasts are obtained, key decision makers should commit themselves to how they will use the information.

The above guidelines had been discussed in academic literature, but were not based on much empirical evidence. Few of these guidelines were used by practitioners in 1960.

Currently, these guidelines do not meet with strong approval. I administered a questionnaire to an interdisciplinary group of academic forecasting experts at a conference in Boulder, Colorado in June 1984. Overall, the 26 respondents reported mild agreement with most of the guidelines, and they did not agree that exponential smoothing was the preferred method that subjective revisions of objective forecasts should be avoided, that causal methods should be used only if highly reliable data were available, or that prior commitment by key decision makers was necessary.

I conducted a literature review to determine whether the empirical evidence obtained over the last 25 years has supported or refuted the guidelines of 1960 and whether additional guidelines are desirable. Fortunately, a great deal of research on forecasting methods has been published since 1960. My search relied heavily upon Armstrong [1985] and Fildes [1981]. In addition, drafts of this paper were circulated to other researchers to elicit additional studies. The findings are provided below in three sections: making forecasts, estimating uncertainty, and gaining acceptance.

Making the Forecasts

The evidence on the six guidelines for making forecasts is summarized in this section.

Decomposition: Decomposition has been widely recommended as a strategy for management science. It can be used to organize an individual's knowledge and also to bring together expertise from different areas. However, little study has been done on its value in forecasting [Armstrong 1985, pp. 57-61]. Based on the meager evidence, the advice to use decomposition seems reasonable. At present, evidence is lacking on how best to decompose problems and on the situations in which decomposition is most useful.

Extrapolation: Much research activity during the past quarter century has concentrated on sophisticated extrapolation methods. (Gardner [1985] provides a recent review of this literature.) However, the empirical research in the area, summarized in the quantitative review in Armstrong [1984] (supported especially by the simulation study by Gardner and Dannenbring [1980] and by the study of 1,001 time series by Makridakis [1982]) suggests there has been negligible gain in accuracy from the use of sophisticated methods (see also McLaughlin [1983] along with other commentaries on the M-Competition). In general, time and money would have been saved had highly M-Competition sophisticated methods been avoided. One exception is that the dampening of the trend term (that is, extrapolating only a portion of the trend) seems to improve accuracy [Gardner and McKenzie 1985].

Expert Opinion: The growth of research on judgmental forecasting has been rapid. The number of publications in this area has been growing at about 14 percent per year over the past 30 years, a crude estimate based on my analysis of references in Kahneman, Slovic, and Tversky [1982]. For reviews of this literature see Armstrong [1985, Chapter 6], Hogarth [1980], and Hogarth and Makridakis [1981]. The research by Kahneman and Tversky (for example, see Kahneman, Slovic, and Tversky [1982]) has been especially influential. It points to numerous shortcomings in the way forecasters use judgment. Sometimes, however, these shortcomings are caused by the questions posed to the forecaster being poorly worded [Tversky and Kahneman 1981; Kruglanski, Friedland, and Farkash 1984]. Some skeptics suggest that psychologists have developed techniques for writing questions that will trick subjects.

Expert opinion is useful in estimating current status, although the direct evidence for this point is sparse. Experts in the subject area often can make adjustments for recent events whose effects have not yet been observed. The research clearly supports the notion that several experts are better than one. Empirical evidence and theoretical arguments by Hogarth [1978] and by Libby and Blashfield [1978] suggest that the optimum number of experts is between five and 20.

More experts should be used where the cost of errors is high, uncertainty is high, the judges have some ability to forecast (yet they differ from one another), and the cost of the experts is low.

Surprising evidence has been obtained on forecasting change. For many years, it was suggested that expert opinion was especially relevant for long-range forecasts. However, research shows it to be less accurate than objective methods in situations where large changes are expected [Armstrong 1985, Chapter 15]. Also surprising is the finding that expertise, beyond a minimal level, is of little value in forecasting change [Armstrong 1985, Chapter 6]. These findings challenge the procedures advocated in 1960. Therefore, forecasters should rely more on objective methods than on expert opinion (especially for long-range forecasts), and they should avoid using expensive experts to forecast change. These findings are counterintuitive, and the reasons behind them are not currently well understood.

Judgmental forecasts may be especially dangerous when dealing with exponential growth. Typically, judges are unable to provide good forecasts of such growth [Wagenaar and Sagaria 1975; Wagenaar and Timmers 1979]. Judgmental forecasts in such situations tend to be conservative. You can test this by asking people to forecast the thickness of a piece of paper if it were folded in half 40 times. Show them the folding step-by-step up to eight folds; then ask them to forecast the final thickness by judgment alone. A group of 20 MBAs in my forecasting course at the University of Hawaii provided the following written predictions: less than one foot: 13; greater than one foot, up to one mile: 5; greater than one mile, up to 2,000 miles: 2. This exercise has been replicated with other groups with similar results. Few estimates exceed one mile. The actual answer is that the folded paper would be thick enough to stretch from the earth to the moon.

If judgmental forecasts must be made for exponential growth problems, two suggestions help: (1) use less data, that is, do not examine the process too frequently (see Wagenaar and Timmers [1978]; and (2) present the historical data as a decreasing function by using the inverse (for example, forecast square miles per person rather than people per square mile [Timmers and Wagenaar 1977]).

Forecasts are often required where data on the dependent variable are absent (for example, new product sales forecasts) or of poor quality (for example, the success of operations to cure lower back pain). Here, the use of bootstrapping is likely to help. Bootstrapping is defined differently by statisticians and computer people, but I use it to mean a "process for translating judgmental forecasting into a set of rules." This can be done by asking forecasters to describe how they make forecasts (either by "protocols," where they describe each step in their thinking as they are making a judgmental forecast, or in some cases they may be able to describe their procedures from memory). More commonly, the rules are inferred by statistical analysis (for example, by regressing a set of judgmental predictions against the information used by the judge).

The research on bootstrapping is substantial (see reviews by Armstrong, [1985] and by Camerer [1981]). Bootstrapping has almost always been more accurate than judgmental forecasts. The gains, although small, have been found in a wide variety of studies involving medicine, psychiatry, production, finance, personnel, and marketing. For repetitive forecasts, bootstrapping is less expensive than judgment. Bootstrapping also provides insight into the way judgmental forecasts are being made, thus offering the possibility of further improvement (for example, it may be that the expert is using invalid relationships). In my opinion, bootstrapping represents the most important advance in forecasting over the past 25 years.

In their study of experts' selections of corporate stocks, Larcker and Lessig [1983] concluded that bootstrapping models developed from memory ("How did you make the forecasts?") were somewhat more accurate than bootstrapping from statistical inference. It appears, however, that this approach is most appropriate when the judges have good insight into their procedures. Neslin [1981], in predicting preferences for a health service, and Schmitt [1978], in a simulated task, each found that statistical inference was superior to asking the: judges to describe their own rules as they made forecasts.

In the engineering and natural sciences literature, bootstrapping research appears under the category of "expert systems," "knowledge-based systems," or "artificial intelligence." As shown in Duda and Shortliffe [1983], much work has been done in this area. In marketing, a similar line of applications is found in the area of "conjoint analysis," where bootstrapping is applied to consumers intentions. A typical application here is to predict sales for new products where the design of the product has many possible variations.

To forecast the effect of a firm's actions, it is often important to forecast the reactions of other major participants in the system. Thus, in considering a major change in marketing strategy, an organization would like to forecast the reactions of its competitors, its own subsequent counteractions, and so on. To make such forecasts, the procedure used in 1960 involved providing the most knowledgeable experts on the topic with the relevant data. It now appears that substantially more accurate forecasts can be made using the outcomes of role-playing (or "gaming") as a basis for the prediction. The evidence, described in Armstrong [1985], is limited to only seven actual situations. However, the superiority of the role-playing predictions was substantial. Subjects made 121 "opinion-based predictions" and were correct in 20 percent of them. In contrast, of the 189 role-playing predictions in these same situations, 70 percent were correct. Furthermore, simple and relatively inexpensive role-playing procedures were able to produce these gains.

The most controversial 1960s recommendation was that forecasts obtained from objective methods should not be revised by expert opinion. This guideline was based on evidence from psychology (for example, Kelly and Fiske [1950]) that expert revisions lessened forecasting accuracy. Some apparently contradictory evidence arose from studies in economic forecasting [Juster 1972; McNees 1975]. Although the issue has not yet been resolved, the results might be explained in this manner: The improvements produced by judgmental adjustments to short-range economic forecasts may be due to improving the estimates of current status; these gains may outweigh the experts deleterious effect on forecasting change. Given the limited evidence to date, it seems appropriate to modify only the estimate of current status. Judgmental inputs to the forecast of change should come prior to the use of objective methods.

Intentions: Research on the use of intentions has not produced many surprises since 1960. Nevertheless, research on this topic has been extensive, and the results useful. The net effect has been substantial, as noted by Perry [1979] in an analysis of errors made in predicting the results of US national political elections (Table 1).

Table 1
An Analysis of Error in Predictions of US Political Elections

Dates	Number of Elections	Mean Absolute Error
1936-48	7	4.0
1950-58	5	1.7
1960-68	5	1.5
1970-178	5	1.0

We can now achieve high response rates from mail and telephone surveys (often in excess of 70 percent), and the impact of the remaining non-response bias can be estimated.

Intentions are useful for short-range predictions of important events given that respondents are willing and able to report their intentions accurately, that they have a plan that they can fulfill, and that new information is unlikely to change the plan. Sewall [1981] showed that intentions provide useful forecasts of catalog mail order sales. His study also indicated that the methods used for scaling the responses, a popular subject for research, have little impact on accuracy.

Causal-Objective Methods: Causal objective methods have not proved useful in all situations. Specifically, they have not been of significant value for short-term forecasting [Armstrong 1985, Chapter 15]. On the other hand, they have not been less accurate than alternative methods. Also, because they add some information, causal-objective methods might be used as one of several approaches to be combined. Furthermore, they are effective for long-range forecasting.

Stepwise regression and other theoretical approaches to the development of causal models have produced few successes and many absurd results. This supports the 1960 guideline stressing that a priori theory should be used. Despite this lack of success, models are often developed without the use of theory, partly because technological advances have reduced the costs of such things as storing data and procedures for exploratory analysis, such as stepwise regression.

Presumably, more observations or data from different sources should yield better estimates of a given causal relationship. This notion is valid, although not so important as originally thought. Multiple sources seem to be important mainly in cases where the direction of the relationship is in doubt.

Good quality data are often thought essential. However, it has been difficult to find evidence that highly reliable data are needed on causal factors. In fact, contradictory evidence exists. For example, unconditional forecasts (where the causal variables must be forecasted) were found to be superior to conditional forecasts in 10 of 12 published papers [Armstrong 1985, p. 241]. This conclusion on quality does not apply, however, to data on the dependent variable [Keren and Newman 1978].

A vast amount of research by econometricians has been devoted to finding good ways to estimate relationships. Dawes and Corrigan [1974], Dorans and Drasgow [1978], and Keren and Newman [1978] suggest that forecast accuracy is not highly sensitive to the estimates of the relationships. The best advice seems to be to pick a small set of reasonable variables, establish the correct direction of the relationship, and obtain rough estimates of the relationships. Some limitations exist for this generalization; and the results of Parker and Srinivasan [1976], and Remus and Jenicke [1978] indicate the need for some caution. It is also likely that good estimates might be important for long-range forecasting. However, far too many resources are devoted to techniques for obtaining better estimates of parameters. Occam's razor should rule: Upon those who advocate more complex procedures rests the burden of proof to show that such procedures are worth the added expense.

Combinations of Forecasts: Academics and practitioners often recommend combining forecasts. This is considered especially useful if the methods are different and if they draw upon different sources of data. Thus, it would be desirable to combine forecasts from extrapolation, judgment, and econometric methods and to use data from independent sources such as consumers, producers, retailers, and experts.

Combinations of forecasts are intuitively appealing. They guard against mistakes, cheating, and unreasonable assumptions and thus help to avoid large errors. They also utilize more information. Nevertheless, surveys of organizations suggest that combined forecasts are used infrequently [Dalrymple, 1986].
Some studies have examined the accuracy of combining forecasts derived from different methods. Falconer and Sivesind [1977] found significant advantages: their combination of an econometric and an extrapolative forecast was 51 percent better than the average of the components and 44 percent better than the best component. However, their econometric model used an ex post forecast. A statistical summary of all of the ex ante forecasts that I could find is presented in Table 2.

Table 2
The percentage error reduction achieved by combining forecasts from different methods: All ex ante forecasts use mean absolute percentage error unless noted otherwise.

Study	Situation	Error Reduction
Port of NY Authority (1957)	Air travel	0.0
Fildes and Fitzgerald (1983)	Balance of payments	1.7[a]
Armstrong, (1985, p. 291)	International photographic market	4.2
Brandon, Fritz and Xander (1983, p. 195)	US GNP	4.8[b,c]
Levine (1960)	Capital expenditures	5.4[b]
Okun (1960)	Housing starts	6.2[b]
Vandome (1963)	National economy (UK)	7.1
O"Herlihy et al. (1967)	Five product categories (UK)	23.4
	Unweighted Average	*6.6*

[a]RMSE (Root Mean Square Error)
[b]different combinations were presented. I used equal weights
[c]all calculations were done using RMSE

Although the sample is small (eight studies), combining is a powerful strategy. The combination of forecasts fro m two different methods yielded a 6.6 percent reduction in error compared to the average component. The error reduction ranged from 0 to 23,percent across these eight studies. Combinations of forecasts from three methods are expected to yield further improvements, but of the two studies available on combining three methods, Rosenzweig [1957] found no gain, and Brandt and Bessler [1983] found an 18 percent error reduction.

Combinations of forecasts derived by similar methods also help. Surprisingly, the gains for combining across extrapolation methods were found to be greater than the results (Table 2). Armstrong [19841 found an average error reduction of 42 percent when two methods were combined (averaged over four studies) while Makridakis and Winkler [1983], in their study of 111 time series, found a 7.2 percent reduction for two methods, with continuing gains as methods were added (for example, a 16.3 percent reduction with five methods). Similar gains have been found when only econometric forecasts were averaged. For example, Brandon, Fritz and Xander [1983] found an 8.7 percent reduction when four econometric forecasts were averaged using equal weights, but this study was based on only four one-quarter-ahead forecasts. Extensive research in judgmental forecasting also shows benefits from combining.

Research to date suggests that equal weighting of the various forecasts is sufficient. A useful approach would be to take an average of the cheapest or more easily understood methods.

Assessing Uncertainty: One function of forecasting is to provide better estimates of the uncertainty associated with a forecast. Uncertainty can have an important influence on decision-making.

Given the increasing sophistication in forecasting, it might be expected that confidence limits would, by now, accompany most forecasts. A reason they do not may be that organizations do not like to discuss uncertainty [March and Simon, 1963, pp. 165-166].

Unfortunately, little research is available on the use of uncertainty estimates in forecasting. One exception is the study of metals forecasts by Rush and Page [1979]. They found that, from 1910 to 1939, 22 percent of the published forecasts contained explicit references to uncertainty. From 1939 to 1964, that figure dropped to 8 percent. The trend is to ignore uncertainty.

One might argue that managers will not understand uncertainty. However, in a survey on weather forecasting, Murphy et al. [1980] found that even the general public understands the meaning of uncertainty.

Unfortunately, uncertainty is typically estimated by judgments and thus is subject to many biases. As illustrated by the following test, people typically underestimate uncertainty:

> Please read the following sentence. "FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF YEARS." Count the number of times the letter F appears in the sentence. Record your answer here ____. Now state your confidence in your answer on a scale from 0 percent, meaning that you are sure you are wrong, to 100 percent, meaning you are sure you are correct. Record your confidence here ____.

Results from a group of executives in a program on forecasting and planning that was held at INSEAD, Fountainebleau, France, in 1981 are shown in Table 3. I have replicated this study many times and obtained similar results.

Table 3
The Letter F Test: Confidence is Not Related to Accuracy

Estimate	Number of Subjects	Average Confidence
Incorrect (3, 4, 5, 7)	12	98
Correct (6)	8	97

As people gain expertise in an area, their tendency toward overconfidence increases [Fischhoff and MacGregor 1982]. 1 suggest that you ignore confidence claims by individuals, except where feedback about outcomes is readily available and easy to summarize. (Believe your weather forecaster!) A useful technique advocated by Benjamin Franklin is to ask experts to summarize arguments against their prediction as well as those in favor; this leads to more realistic assessments of uncertainty [Koriat, Lichtenstein, and Fischhoff 1980].

Objective (quantitative) approaches should be used to estimate uncertainty whenever possible. To assess the uncertainty for a one-year forecast of automobile sales, for example, one should use the accuracy of previous one-year forecasts of automobile sales.

Prior Commitment: Forecasters should obtain specific prior commitment from decision makers as to how they will use the forecasts. Without prior commitment, decision makers are unlikely to be influenced by forecasts that conflict with their expectations. This was shown in a study by Griffith and Wellman [1979]; although the objective forecasts by an outside consultant of hospital bed needs were more accurate than the subjective forecasts by management' management followed their own forecasts. Baker [1979] showed that hurricane forecasts were frequently ignored when it would have seemed rational to pay attention to them.

How then can the forecaster gain prior commitment? Scenarios may help. By casting a potential forecast into a story (the scenario), decision makers are much more likely to find the forecast plausible. Its perceived likelihood can be increased by using the following techniques:

(1) Use concrete examples [Read 1983; Anderson 1983b].
(2) Make the description vivid [Hamill, Wilson, and Nisbett 1980].
(3) Include events that will seem representative of the situation, even if it means using irrelevant events [Tversky and Kahneman 1982].
(4) Include events that will make it easy for the decision makers to recall supporting evidence (see Tversky and- Kahneman [1973] on "availability").
(5) Link the events to show causality in a highly plausible manner [Kahneman and Tversky 1979].
(6) Include commensurate measures across alternative scenarios so that the relative impact can be manipulated; this is effective even if the commensurate measures are irrelevant [Slovic and McPhillamy 1974].
(7) Ask the decision makers to project themselves into the situation [Anderson 1983a].

(8) Ask the decision makers to predict how they would act if they were participants in the scenarios [Sherman 1980].

Note the danger, however: These techniques could seriously distort judgmental forecasts [Gregory, Cialdini, and Carpenter 1982]. In other words, it would be easy to "lie with scenarios."

Conclusions

Some of the old-fashioned forecasting principles have held up well over the past 25 years. The research supports many of the guidelines (especially for decomposition, extrapolation, intentions, combining forecasts, uncertainty, and prior commitment by decision makers). However, it does suggest more effective ways to use these guidelines. Two important examples are

?? Various techniques for scenario writing have been found effective in gaining prior commitment, and
?? The use of combined forecasts was shown to be highly effective. Rather than weight by confidence, however, it appears less expensive and safer to weight forecasts from each of the reasonable methods equally.
?? Significant advances were achieved i n judgmental forecasting. The following guidelines are new:
?? Judgment provides a poor approach to long-range forecasting.
?? High expertise in the area to be forecast is of little value in forecasting change.
?? Role-playing provides more accurate forecasts in conflict situations than unaided opinions.
?? Bootstrapping can provide more accurate forecasts than judgment, and it is often less expensive.
?? Extensive study in two areas yielded discouraging results:
?? The research on sophisticated approaches to objective methods has had only modest payoff.
?? The findings on causal models have been surprising. It now appears that crude estimates of causal relationships are sufficient, with the possible exception of those situations involving large changes.

This review of the literature implies that the research budget on forecasting is being misallocated. Table 4 summarizes the eight areas for research and suggests changes for the future. The last column lists my opinions on the potential payoff for research on each topic. The amount spent on sophisticated approaches for extrapolation or for causal models should be reduced. Research on judgmental forecasting should receive an increased budget. Additional research is also needed on methods for gaining acceptance of forecasts and on role-playing.

Table 4
Progress in Research on Forecasting
Research from 1960 to 1984

Topic	Effort	Payoff	Examples of Conclusions	Future Research Prospects
Decomposition	Little	Low	Helpful where uncertainty is high	Good
Extrapolation	Very Great	Low	Few differences in accuracy among methods Trends should be dampened	Modest
Expert Opinion	Great	Very High	Limited value of expertise in forecasting large changes Role playing accurate in conflict situation Bootstrapping more accurate	Excellent
Intentions	Modest	Modest	Ways to reduce response and non-response bias	Modest
Causal Methods	Very Great	Low	Simplicity is a virtue	Modest
Combined Forecasts	Little	High	Combinations yield substantial gains	Modest
Uncertainty	Little	Modest	Overconfidence in judgmental estimates Argue against your forecast	Good
Implementation	Little	Modest	Scenarios can help to gain prior commitment	Modest

This has been primarily a qualitative review of research on forecasting. As the research grows, opportunities arise using "meta-analysis," the quantitative assessment of previous studies. In other words, hypotheses about forecasting can be assessed by counting the number of studies for or against each hypothesis. Statistical tests can be made using each study as an observation. Procedures for meta-analysis are described in Glass, McGaw and Smith [1981] and in Hunter, Schmidt and Jackson [1982]. In this paper, I used meta-analysis in an attempt to assess the value of combining forecasts (Table 2). In Armstrong [1984], I provided another example, a statistical analysis of forecast accuracy in studies that compared simple and complex extrapolation methods. Efforts to conduct meta-analysis in forecasting are aided by Fildes' bibliographies (1981, 1983), the references in Armstrong [1985], the Social Science Citation Index and the many computer search files now available.

When Lee [1980] reviewed the major advances that had resulted from research on management, he researched some dismal conclusions: Many important techniques have been known for decades, even centuries, while many recent "advances" have proved to be of little value. In contrast, research in forecasting shows substantial gains over the past quarter century.

Acknowledgments

Lawrence D. Brown, Derek Bunn, Robert Fildes, Lisa Krutzel, Herman Stekler, and Steven C. Wheelwright provided helpful comments on an early version of this paper.

References

Anderson, Craig A. 1983a, "Imagination and expectation: The effect of imagining behavioral scripts on personal intentions," *Journal of Personality and Social Psychology* Vol. 45, No. 2, pp. 293-305.

Anderson, Craig A. 1983b, "Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs," *Journal of Experimental Social Psychology*, Vol. 19, No. 2, pp. 93-108.

Armstrong, J. Scott 1984, "Forecasting by extrapolation: Conclusions from 25 years of research," *lnterfaces* Vol. 14, No. 6, pp. 52-61.

Armstrong, J. Scott 1985, *Long-Range Forecasting: From Crystal Ball to Computer*. (second edition) John Wiley and Sons, New York.

Baker, Earl Jay 1979, "Predicting response to hurricane warnings: A reanalysis of data from four studies," *Mass Emergencies* Vol. 4, No. 1, pp. 9-24.

Brandon, Charles; Fritz, Richard; and Xander, James 1983, "Econometric forecasts: Evaluation and revision," *Applied Economics* Vol. 15, No. 2, pp. 187-201.

Brandt, Jon A. and Bessler, David A. 1983, "Price forecasting and evaluation: An application in agriculture," *Journal of Forecasting*, Vol. 2, No. 3, pp. 237-248.

Camerer, Colin 1981, "General conditions for the success of bootstrapping models," *Organizational Behavior and Human Performance*, Vol. 27, No. 3 (June), pp. 411-22.

Dalrymple, Douglas J. 1986, "Sales forecasting practices: Results from a U.S. survey," *International Journal of Forecasting* 3, 379-391.

Dawes, Robyn M. and Corrigan, Bernard 1974, "Linear models in decision making," *Psychological Bulletin*, Vol. 81, No. 2, pp. 95-106.

Dorans, Neil and Drasgow, Fritz 1978, "Alternative weighting schemes for linear prediction," *Organizational Behavior and Human Performance*, Vol. 21, No. 3 (June), pp. 316-345.

Duda, Richard O. and Shortliffe, Edward H. 1983, "Expert systems research," *Science*, Vol. 220, No. 4594 (April 15), pp. 261-268.

Falconer, Robert and Sivesind, Charles M. 1977, "Dealing with conflicting forecasts: The eclectic advantage," *Business Economics*, Vol.12, No.4, pp.5-11.

Fildes, Robert (with Dews, David and Howell, Syd) 1981, *A Bibliography of Business and Economic Forecasting*. Gower Publishing, Farnborough, Hants, England: (see also Supplement 1, 1979-1981, available from the Manchester Business School, Working Paper Series 87, 1984).

Fildes, Robert and Fitzgerald, M. Desmond 1983, "The use of information in balance of payments forecasting," *Economica*, Vol. 50, No. 199, pp. 249-258.

Fischhoff, Baruch and MacGregor, Don 1982, "Subjective confidence in forecasts," *Journal of Forecasting*, Vol. 1, No. 2, pp. l55-172.

Gardner, Everette S., Jr., 1985 "Exponential smoothing: The state of the art," *Journal of Forecasting*, Vol. 4, No. 1, pp. 1-38.

Gardner, Everette S., Jr. and McKenzie, E. 1985, "Forecasting trends in time series," *Management Science*, Vol. 31, No. 10 (October), pp. 1237-1246.

Gardner, Everette S., Jr. and Dannenbring, David G. 1980, "Forecasting with exponential smoothing: Some guidelines for model selection," *Decision Sciences*, Vol. 11, No. 2 (April), pp. 370-383.

Glass, Gene V.; McGaw, Barry; and Smith, Mary L. 1981, *Meta-Analysis in Social Research*. Sage Publications, Beverly Hills, California.

Gregory, W. Larry; Cialdini, Robert B.; and Carpenter, Kathleen M. 1982, "Self relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so?" *Journal of Personality and Social Psychology*, Vol. 43, No. 1, pp. 89-99.

Griffith, John R. and Wellman, B. T. 1979, "Forecasting bed needs and recommending facilities plans for community hospitals: A review of past performance," *Medical Care*, Vol. 17, No. 3, pp. 293-303.

Hamill, Ruth; Wilson, T; D.; and Nisbett, R. E. 1980, "Insensitivity to sample bias: Generalizing from atypical cases," *Journal of Personality and Social Psychology*, Vol. 39, No. 4, pp. 578-589.

Hogarth, Robin M. 1980, *Judgment and Choice: The Psychology of Decision*, John Wiley and Sons, New York.

Hogarth, Robin M. 1978, "A note on aggregating opinions," *Organizational Behavior and Human Performance*, Vol. 21, No. I (February), pp. 4-46.

Hogarth, Robin M. and Makridakis, Spyros 1981, "Forecasting and planning: An evaluation," *Management Science*, Vol. 27, No. 2, pp. 115-138.

Hunter, John E.; Schmidt, Frank L.; and Jackson, Gregg B. 1982, *Meta-Analysis: Cumulating Research Findings Across Studies*, Sage Publications, Beverly Hills, California.

Juster, F. Thomas 1972, "An evaluation of the recent record in short-term forecasting," *Business Economics*, Vol. 7, No. 3 (May), pp. 22-26.

Kahneman, Daniel; Slovic, Paul; and Tversky, Amos 1982, *1udgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, England.

Kahneman, Daniel and Tversky, Amos 1979, "The simulation heuristic," Reprinted in Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.) 1982, *1udgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, England, pp. 201208.

Kelly, E. Lowell and Fiske, D. W. 1950, "The prediction of success in the VA training program in clinical psychology," *American Psychologist*, Vol. 5, No. 8, pp. 395 406.

Keren, Gideon and Newman, J. Robert 1978, "Additional considerations with regard to multiple regression and equal weighting," *Organizational Behavior and Human Performance*, Vol. 22, No. 2 (October), pp. 143-164.

Koriat, Asher; Lichtenstein, Sarah; and Fischoff, Baruch 1980, "Reasons for confidence," *Journal of Experimental Psychology: Hunzan Learning and Mnnory*, Vol. 6, No. 2, pp. 107-118.

Kruglanski, Arie W.; Friedland, Nehemia; and Farkash, Ettie 1984, "Lay persons' sensitivity to statistical information: The case of high perceived applicability," *Journal of Personality and Social Psychology*, Vol. 46, No. 3, pp. 503-518.

Larcker, David F. and Lessig, V. Parker 1983, "An examination of the linear and retrospective process tracing approaches to judgment modeling," *Accounting Review*, Vol. 58, No. 1, pp. 58-77.

Lee, James A. 1980, *The Gold and the Garbage in Management Theories and Prescriptions*, Ohio University Press, Athens, Ohio.

Levine, Robert A. 1960, "Capital expenditures forecasts by individual firms," in *The Quality and Significance of Anticipations Data*, National Bureau of Economics Research, Princeton, New Jersey, pp. 351-366.

Libby, Robert and Blashfield, Roger K. 1978, "Performance of a composite as a function of the number of judges," *Organizational Behavior and Human Performance*, Vol. 21, No. 2 (April), pp. 121-129.

Makridakis, Spyros et al. 1982, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, Vol. 1, No. 2, pp. 111-153.

Makridakis, Spyros and Winkler, Robert L. 1983, "Averages of forecasts: Some empirical results," *Management Science*, Vol. 29, No. 9, pp. 987-996.

March, lames G. and Simon, Herbert A. 1963, *Organizations*, John Wiley and Sons, New York.

McLaughlin, Robert L. 1983, "Forecasting models: Sophisticated or naive?" *Journal of Forecasting*, Vol. 2, No. 3, pp. 27-276.

McNees, Stephen K. 1975, "An evaluation of economic forecasts," *New England Economic Review* (November-December), pp. 3-39.

Murphy, Allan H.; Lichtenstein, S.; Fischhoff, B.; and Winkler, R. L. 1980, "Misinterpretations of precipitation probability forecasts," *Bulletin of the American Meteorological Society*, Vol. 61, No. 7, pp. 695-701.

Neslin, Scott A. 1981, "Linking product features to perceptions: Self-stated versus statistical]y revealed importance weights," *Journal of Marketing Research*, Vol. 18, No. 1, pp. 80-6.

O'Herlihy C. et al. 1967, "Long-term forecasts of demand for cars, selected consumer durables and energy," *National Institute Economic Review*, Vol. 40, No. 40 (May), pp. 34 61.

Okun, Arthur H. 1960, '`The value of anticipations data in forecasting national product," in *The Quality and Economic Significance of Anticipations Data*. National Bureau of Economic Research, Princeton, New Jersey.

Parker, Bamett R. and Srinivasan, V. 1976, "A consumer preference approach to the planning of rural primary care facilities," *Operations Research*, Vol. 24, No. 5, pp. 991-1025

Perry, Paul 1979, "Certain problems in election survey methodology," *Public Opinion Quarterly*, Vol. 43, pp. 312-325.

Port of New York Authority 1957, *Air Travel Forecasting 1965-1975*. Eno Foundation, Saugatuck, Connecticut.

Read, Stephen J. 1983, "Once is enough: Causal reasoning from a single instance," *Journal of Personality and Social Psychology*, Vol. 45, No. 2, pp. 323-334.

Remus, William E. and Jenicke, Lawrence O. 1978, "Unit and random linear models in decision making," *Multivariate Behavioral Research*, Vol. 13, No. 1, pp. 215-221.

Rosenzweig, James E. 1957, *The Demand for Aluminum: A Case Study in Business Forecasting*. University of Illinois, Bureau of Economics and Business Research, Champaign, Illinois.

Rush, Howard and Page, William 1979, "Long-term metals forecasting: The track record: 1910-1964," *Futures*, Vol. 11, No. 4 (August), pp. 321-33

Schmitt, Neal 1978, "Comparison of subjective and objective weighting strategies in changing task situations," *Organizational Behavior and Human Performance*, Vol. 21, No. 2 (April), pp. 171-188.

Sewall, Murphy A. 1981, "Relative information contributions of consumer purchase intentions and management judgment as explanators of sales," *Journal of Marketing Research*, Vol. 18, No. 2, pp. 249-253.

Sherman, Steven J. 1980, "On the self-erasing nature of errors of prediction," *Journal of Personality and Social Psychology*, Vol. 39, No. 2, pp. 211-221.

Slovic, Paul and McPhillamy, Douglas 1974, "Dimensional commensurability and cue utilization in comparable judgment," *Organizational Behavior and Human Performance*, Vol. 11, No. 2 (April), pp. 172-194.

Timmers, Han and Wagenaar, Willem A. 1977, "Inverse statistics and misperception of exponential growth," *Perception and Psychophysics*, Vol. 21, No. 6, pp. 558-562.

Tversky, Amos and Kahneman, Daniel 1973, "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, Vol. 5, No. 2, pp. 207-232.

Tversky, Amos and Kahneman, Daniel 1981, "The framing of decisions and the psychology of choice," *Science*, Vol. 211, No. 4481, pp. 453-458.

Tversky, Amos and Kahneman, Daniel 1982, "Judgments of and by representativeness," in Daniel Kahneman, Paul Slovic and Amos Tversky (eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, pp. 84-98.

Vandome, Peter 1963, "Econometric forecasting for the United Kingdom," *Bulletin of the Oxford University Institute of Economics and Statistics*, Vol. 25, No. 4, pp. 239-281.

Wagenaar, Willem A. and Sagaria, Sabato D. 1975, "Misperception of exponential growth," *Perception and Psychophysics*, Vol. 18, No. 6, pp. 416-22.

Wagenaar, Willem A. and Timmers, Han 1978, "Extrapolation of exponential time series is not enhanced by having more data points," *Perception and Psychophysics*, Vol. 24, No. 2, pp. 182-184.

Wagenaar, Willem A. and Timmers, Han 1979, "The pond-and-duckweed problem: Three experiments in the misperception of exponential growth," *Acta Psychologica*, Vol. 43, No. 3, pp. 23-251.