Published in V.H. Dale and M.E. English, eds., *Tools to Aid Environmental Decision Making*, 1999, New York: Springer-Verlag, pp. 192-225.

Forecasting for Environmental Decision Making

J. Scott Armstrong
The Wharton School, University of Pennsylvania

> "The Ford engineering staff, although mindful that automobile engines provide exhaust gases, feels that these waste vapors are dissipated in the atmosphere quickly and do not present an air pollution problem." Official spokesperson for the Ford Motor Company in 1913 in response to a letter from the Los Angeles county supervisor.
>
> Cerf and Navasky, 1991, p. 39.

Those making environmental decisions must not only characterize the present, they must also forecast the future. They must do so for at least two reasons. First, if a no-action alternative is pursued, they must consider whether current trends will be favorable or unfavorable in the future. Second, if an intervention is pursued instead, they must evaluate both its probable success given future trends and its impacts on the human and natural environment. Forecasting, by which I mean *explicit processes* for determining what is likely to happen in the future, can help address each of these areas.

Certain characteristics affect the selection and use of forecasting methods. First, the concerns of environmental forecasting are often long term, which means that large changes are likely. Second, environmental trends sometimes interact with one another and lead to new concerns. And third. interventions can also lead to unintended changes.

This chapter discusses forecasting methods that are relevant to environmental decision making, suggests when they are useful, describes evidence on the efficacy of each method, and provides references so readers can get details about the methods. A key consideration is whether or not the forecasting methods are designed to assess the outcomes of interventions. The chapter then examines issues related to presenting forecasts effectively. Finally, it describes an audit procedure for determining whether the most appropriate forecasting tools are being used.

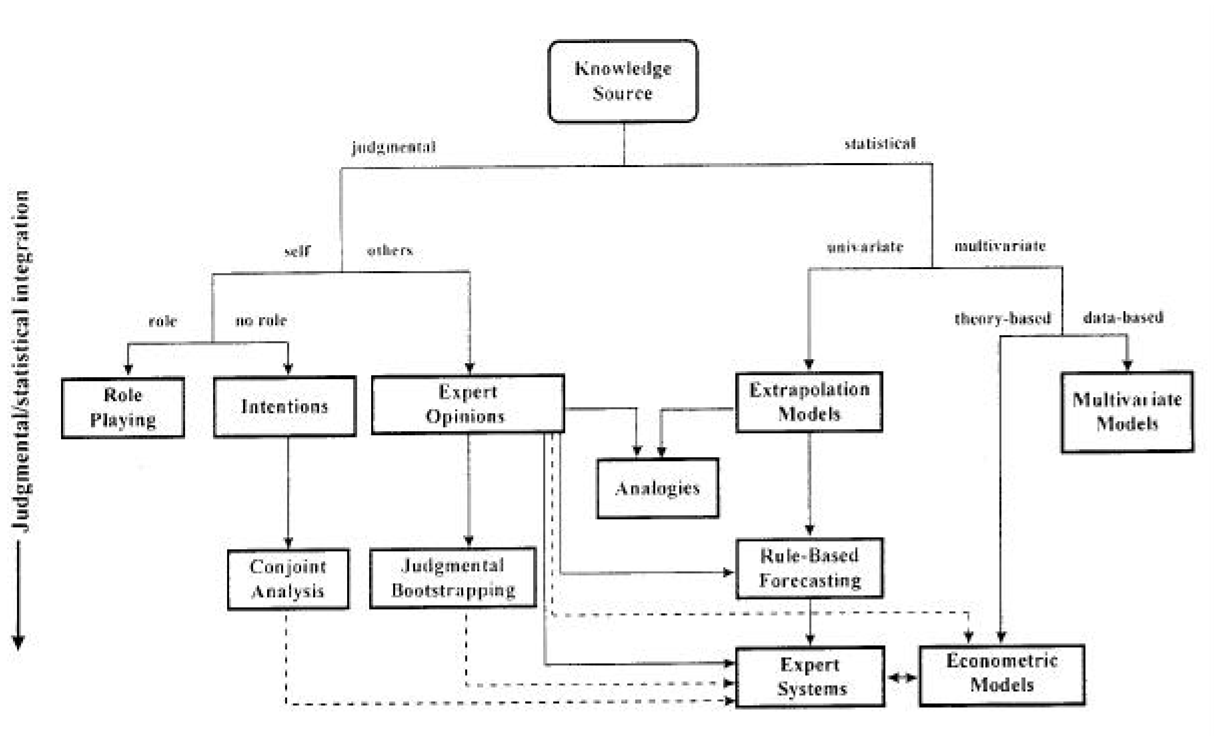
A Framework for Forecasting

Figure 1 shows possible forecasting methods and how they relate to one another. The figure is designed to represent all approaches to forecasting. The methods are organized according to the types of knowledge. As one moves down the chart, the integration of statistical and judgmental methods increases.

Judgmental methods are split into those involving one's own behavior in given situations (intentions or role playing) and expert opinions. The intentions method asks people how they would act in a given situation. Role playing examines how people act in a situation where their actions arc influenced by a role. Experts can be asked to make predictions about how others will act in given situations. They can also identify analogous situations, and forecasts can be based on extrapolations from those situations.

Intentions and expert opinions can be quantified by relating their "predictions" to various causal factors with, for example, regression analysis. Expectations about one's own behavior are referred to as conjoint analysis (e.g., given alternatives having a bundle of features that have been varied according to an experimental plan). Expert opinions about the behavior of others (which can also be based on an experimental design, but are often based on actual data) are referred to as judgmental bootstrapping.

Figure 1
Characteristics of forecasting methods and their relationships (dotted lines represent possible relationships)



The statistical side of the methodology tree has univariate and multivariate branches. The univariate branch leads to extrapolation methods. By drawing upon expert opinions, one can develop rule-based forecasting. This procedure uses domain knowledge to select and weight extrapolation methods.

Expert systems use the rules of experts. These rules might be based on protocol sessions in which experts are asked to describe how they make forecasts while they are actually in the process of forecasting. Alternatively, experts' rules could be formalized by drawing upon estimates produced by judgmental-bootstrapping models. Quite commonly. developers of expert systems also draw upon empirical studies of relationships, with some of those studies involving econometric models. Another possible source of information is embodied in relationships estimated by conjoint analysis.

The multivariate branch is split into data-based and theory-based branches. In the theory-based approach, the analyst formulates a model and then refines the parameter estimates based on information gleaned from experts and data. These constructs are referred to as econometric models. Data-based approaches try to infer relationships from the data. I refer to them as multivariate models.

In all, then, 11 approaches to forecasting are proposed. More attention will be given here to those for which there is stronger evidence. For example, despite an immense amount of research effort, the evidence that multivariate methods provide any benefits to forecasting is weak. The case for expert systems is stronger, but I am not aware of studies that assess the use of expert systems for producing environmental forecasts.

Forecasts Without Interventions

Measurement procedures for environmental decision making are improving, and more relevant data are being collected. As a result, environmental decision makers have increasing amounts of information about trends in the environment and related socioeconomic, political, and legal conditions. Some of these trends. such as increases in pollution, may seem unfavorable; whereas other trends, such as advances in technology, may seem favorable.

Whether or not a trend is favorable depends upon one's point of view. For example, some observers fear the effects of global warming while others believe that its effects would be beneficial. In structuring the forecasting problem, then, it is important to forecast the trends and their effects on all affected populations, human and nonhuman. For example, what are the predicted effects of increasing amounts of pollution on residents, land owners, people living downwind or downstream, product manufacturers, consumers, waste-disposal firms, and wildlife?

Forecasts with Interventions

The primary reasons for making explicit forecasts arc to determine whether to intervene, and if so, how. Forecasting can help decision makers to assess alternative interventions.

Interventions can affect many aspects of the system. For example, to reduce pollution in Santiago, Chile, in the early 1990s, the government restricted the use of automobiles in the central business district by allowing entry to only those cars whose license plates ended in even numbers on one day and those ending with odd numbers on the next day. Such a plan might have an impact on those who sell automobiles (perhaps refurbishing old cars so people can have cars with both even- and odd-numbered license plates), commuters (perhaps becoming less productive because they spend more time on public transit), and so forth. Forecasters need to consider the effects on each group and how they will react to an intervention.

A forecast of the effects on only part of the system might be worse than no forecast at all because it might lead to unwise decisions. For exa mple, the concern over disposal problems with plastic packaging put so much public pressure on McDonald's that they switched to paper packaging. Some analysts have concluded that paper packaging not only is less convenient for workers and customers, but also creates more pollution when one considers the entire cycle from producing paper packaging, using it, and disposing of it.

Use of Forecasts

Plans are often confused with forecasts. Plans are sets of actions to help deal with the future. Forecasting (or predicting) is concerned with determining what the future will be. A plan is an input to the forecasting model. If the forecasts arc undesirable, then one might change the plan, which, in turn, could change the forecast. The point to remember is that good plans depend on good forecasts.

In practice, forecasts are sometimes used to motivate people. More properly, people should be motivated by plans (e.g., "meet this plan. and we will pay you a bonus of 25 percent").

Decisions have often been made, before any formal forecasting has been done. In such cases, the forecast serves little purpose other than to annoy people if it conflicts with their decision or to please them if it supports their decision. For the forecast to be used effectively, it should be prepared before decisions arc made.

Not only the expected outcome, but also other likely outcomes (such as the best and worst outcomes) should be forecast. If the worst outcome poses too much risk, forecasts should be made for alternative interventions.

Methods for Environmental Forecasting

Environmental forecasting often involves decisions that have long-term consequences. In addition, these forecasts are likely to be subject to severe biases, depending on one's perspective. To address these issues, I suggest four principles:

- ? Use relevant information
- ? Ensure that the information and procedures are objective
- ? Structure the inputs to the forecasting procedure
- ? Use methods that arc no more complex than necessary

Relevant information. Few people would argue that obtaining more information is not useful, but in some situations, more information does not improve forecast accuracy. If the additional information is irrelevant, it is likely to reduce forecast accuracy. For example, in many professions, the height of people is not important to job performance; however, height plays a role in job interviews when people make predictions about a candidate's ability to do a job. The danger of using irrelevant information is especially serious when using judgmental procedures. For example, knowing that a celebrity has taken a strong stand on an environmental issue might detract from one's ability to make a good judgment on the issue: more than likely, the person's celebrity status has nothing to do with his or her making an informed judgment. Also, more information is not likely to improve fore casting accuracy if it is used in an unstructured manner (Armstrong. 1985, pp. 100-102).

Objective information and procedures. Before data collection and analysis begin, decision makers should agree on what information is relevant and how it should be analyzed. These decisions should be made by people who are not biased to any particular viewpoint, at least in that they will not personally gain. It would seem sensible here to draw upon those with expertise in forecasting methods.

To avoid collecting data that might be biased, one should seek alternative sources of data. Brenner et al. (1996) concluded that subjects who received biased data were less accurate in their predictions but more confident, even when it was painfully obvious that the data were biased and incomplete.

Structured inputs. Forecasting can be improved by structuring the problem to make efficient use of available information. One of the most useful tools for structuring forecasting problems is decomposition. The safest way to decompose problems is to break them into additive elements. For example, population is often decomposed to forecast births, deaths, immigra tion, and emigration; these separate forecasts arc then added to form a composite forecast of population. Alternatively, multiplicative decomposition is sometimes useful. For example, certain air pollutants might interact, so it would not be sufficient to merely add their effects. But multiplicative decomposition can be risky because the errors in each element get multiplied by those in all the other elements. MacGregor (1999) summarized research on this topic and concluded that multiplicative decomposition improves accuracy when uncertainty is high and extreme numbers (large or small) are involved. For example, it would be difficult to make judgmental estimates on the number of pounds of harmful automobile pollution produced each year in the United States. Typically, one can decompose problems to avoid forecasting extreme numbers. However, multiplicative decomposition is sometimes detrimental when numbers arc not large and uncertainty is not great.

Simple methods. One of the more interesting findings from empirical research on forecasting is that relatively simple methods provide forecasts that are as accurate as those from more complex methods. It follows, then, that forecasting methods should be no more complex than necessary. This conclusion seems to cover a wide range of conditions. Simple methods reduce costs and they aid understanding among forecasters and decision makers. They also reduce the likelihood of mistakes. Mistakes do occur, however, even for relatively simple methods, such as exponential smoothing. Gardner (1984) cited 23 books and articles containing errors in model formulations for exponential smoothing. The more complex a process is, the more likely it is that a mistake might creep in and remain undetected.

The use of simple methods should he welcome to decision makers. Yokum and Armstrong's (1995) survey of forecasters concluded that ease of understanding, implementation, and use were almost as important as accuracy when selecting forecasting methods.

One way to avoid imposing one's own bias is to seek forecasts that have been provided by independent third parties. This procedure might also save money. Before using a published forecast, you should have details about the source of the forecast along with the methods and data that were used. Unfortunately. these details are often omitted.

Our major concern in this chapter is with the most useful methods for environmental forecasting. Most approaches from Figure 1 are relevant. However, data-based multivariate forecasting models are inappropriate because they ignore the substantial body of expertise that is typically available on environmental issues. In addition. they arc expensive, difficult to communicate, complex, subject to the imposition of biases or mistakes that are hard to detect. and often misunderstood, even by leading proponents. Finally, despite enormous efforts in their development, little testing has been done of their predictive validity, and what has been done suggests that this approach is not promising.

Below are descriptions of the methods that seem most relevant given the guiding principles for environmental forecasting Table.1 lists these methods along with brief descriptions of their uses, advantages, and disadvantages.

Table 1. Methods for environmental forecasting.

Tools	Descriptions	Advantages	Disadvantages
		Judgment	
Role playing	Simulates the interaction of conflicting groups in a decision process	Taps the experience and knowledge of decision makers in a realistic interaction	Group dynamics may influence players in ways that are unintended; selected players may not be knowledgeable
Intentions	Determines what actions decision makers would take in certain circumstances	Provides the perspective of those who will actually make decisions	Subject to sampling and questioner bias; changes with time as other factors come into play
Expert opinion (Delphi)	Experts forecast how others will behave	Information is derived from knowledgeable sources; inexpensive to perform; useful even when data are poor or lacking	Overly influenced by the current situation
Analogies	Examines how similar situations have turned out	Based on real world experiences	May have poor correspond-ence to current situation
Conjoint analysis	To gauge citizens' reactions to aspects of an intervention	Citizens often have a good sense of how they will respond	Expensive
Judgmental bootstrapping	To gauge citizens' reactions to aspects of an intervention	Experts often have useful information; inexpensive	Experts may lack relevant knowledge
		Extrapolation	
Exponential smoothing	Extends historical values into the future	Reliable; reproducible; simple methods produce accurate results; limits introduction of bias	Inaccurate, given discon-tinuities or unstable trends; ignores domain knowledge; especially risky when trends are contrary to expectations
		Econometric Methods	
Single-equation, theory-based models	Forecast based on causal relationships	Results are firmly grounded in domain knowledge and theory; especially useful when large changes occur; alternative interventions can be compared; can aid in the construction of confidence intervals	Complex procedures not easily understood by decision makers; may lack data on causal variables; may overlook key variables; expensive
		Integrated Forecasts	
Rule-based forecasting	Assign differential weights to extrapolative forecasts	Based on published research and expert advice about forecasting methods; improves accuracy; offers protection against large errors; aids objectivity; incorporates manager's knowledge	Added complexity and cost
Expert systems	Apply rules determined by experts and by empirical studies	Formalizes available knowledge about a situation	Expensive; little information about forecast validity

Tools	Descriptions	Advantages	Disadvantages
		Combined Forecasts	
Equal weights	Combine forecasts from several methods, giving the results of each method the same influence on the final result	Improves accuracy; offers protection against huge errors or mistakes; aids objectivity; useful given uncertainty about which method is best	Ignores domain knowledge; not appropriate when better method is known

Judgmental Methods

Judgmental forecasting involves methods that process information by experts, rather than by quantitative methods. The experts might have access to data, and their approach might be structured, but the final forecasts arc the result of some process that goes on in their heads.

Before discussing tools that aid judgmental forecasting, it is important to mention one tool that is widely used and well accepted, but which typically harms accuracy and leads to an unwarranted gain in confidence. The culprit is the traditional (unstructured) group meeting. Besides the biases inherent in unstructured meetings (such as the influence of the boss), the group's information is likely to be poorly used (Armstrong, 1985, pp. 1211-121).

Judgmental forecasts are susceptible to various biases. To reduce biases, one should select unbiased experts (i.e., those who have nothing to gain from a forecast that is either too high or too low). In addition, care should be given to how the forecasting problem is formulated. Questions should be structured to use the judges' knowledge most effectively, pretested to ensure that the experts understand them, and worded in different ways to see if that affects the forecasts. Such procedures are particularly important when forecasting sensitive issues, such as the effects of global warming.

The use of structured procedures can greatly improve the accuracy of judgmental forecasts. Structure is easy to apply and involves only modest costs. I discuss four structured judgmental procedures that should be of interest for environmental forecasting: (1) role playing, which uses subjects to act out relevant interactions to determine what they would do when affected by an intervention; (2) intention surveys, which use statements by key participants in the system about what they expect to do given certain trends or interventions; (3) Delphi. which uses expert judgment to forecast trends or the effects of intervention: and (4) analogies, where experts try to generalize from similar situations. Brief attention is given to conjoint analysis and to judgmental bootstrapping.

Role Playing

Role playing involves asking subjects to adopt the viewpoints of groups in a negotiation situation and having them act out the interactions. When the interactions of conflicting groups are important to the outcome, role playing provides a way to simulate this interaction. If new and important interventions would lead to behaviors that are dependent upon the interactions among decision makers, then role playing is likely to be more relevant than intentions. With intentions, decision makers would have to predict what they would do initially, how they would modify their decisions in reaction to the decisions made by others, how others would respond to this reaction, and so on. This chain of events is often too complex for the respondent, so it makes sense to act it out.

To use role playing to forecast the outcome of an intervention, such as a tax on air pollution, one would write short descriptions of the problem and of the roles of key decision makers. Different materials can be prepared to test alternative interventions. These guidelines should be lfllowcd:

- Use props to make the situation realistic.
- Select subjects who can act the role (interestingly, the selection of subjects does not seem to he a critical aspect for the accuracy of role playing).
- Subjects should receive their roles before they receive any information about the situation, and they should not step out of their roles.
- Subjects should act as they would act if they were actually in such a role.

- Subjects should improvise as needed.

Forecasts would be based on the outcomes of the role-playing sessions. Ideally, possible outcomes can be identified in advance. However, if the range of possible outcomes is uncertain, one should leave the materials open-ended and ask research assistants to code the outcome, of the role-playing sessions. If the session does not lead to an outcome, one can ask the players to predict what would have happened had it continued to a conclusion.

Prediction intervals can be constructed by assessing the proportion of times than a certain outcome occurs in a set of role-playing sessions. The standard error of this estimate can then be obtained by using the formula for the standard error of a proportion, with the number of role-playing sessions as the sample size. Prediction intervals would be expected to be larger than this estimate because of possible response biases.

While role playing has been used as a predictive device in the military and in the legal profession for many years, research on its value as a predictive technique is limited. Armstrong (1987) and Armstrong and Hutcherson (1989) report on studies that compared unaided opinions with role playing for eight situations. These included the conflict between Mexico and the United States, which led to the United States acquiring Texas; the marketing of Upjohn's drug, Panalba, after a commission concluded that it was causing unnecessary deaths; the presidential-election conventions held by the party that was out of power; an attempt by Philco to gain the agreement of supermarket owners to allow them sell appliances in supermarkets; negotiations between the National Football League players and the owners in 1982; an attempt by artists in The Netherlands to have the government buy their artwork if they could not find anyone to purchase it; a negotiation over the royalties for an academic journal; and whether bombing North Vietnam would be a good strategy for the United States in the 1970s. Role playing was superior to opinions on seven of the tests; it tied on one (that involving political conventions). Averaging across the eight situations, which involved 226 role-playing sessions, role playing was correct for 63.6 percent of the cases versus only 18.2 percent for unaided opinions. A listing of these experimental comparisons is provided in Table 2.

Table 2. Role playing versus opinions.

Situation	Conflict among	Percent correct (Sample size) Opinion	Percent correct (Sample size) Role play
US-Mexico	Countries	1 (1)	57 (96)
*Panalba (prescription drug)	Stockholder and consumer	34 (63)	79 (57)
US political convention	Candidates	67 (12)	67 (12)
*Philco appliances	Manufacturer and retailer	3 (37)	75 (12)
*NFL football	Players and owners	27 (15)	60 (10)
*Artists in Holland	Artists and government	3 (31)	29 (14)
*Journal royalties	Publisher and editors	12 (25)	42 (24)
North Vietnam bombing	Countries	0 (1)	199 (1)
	Averages	18.2	63.6

* Based on low-fidelity role playing.

Although none of the eight validation situations involved environmental decisions, they all involved conflicts between groups. Thus, I would expect that role playing would he useful for forecasts involving environmental conflicts, such as whether to charge farmers more for cattle-grazing rights on government lands or what restrictions would be effective to control certain types of fishing. Based on research to date, role playing would seem to be more accurate than other methods (except experimentation) for forecasts involving environmental conflicts.

Role playing. however, is inexpensive relative to experimentation. Many of the studies were conducted using role playing sessions that lasted less than an hour. In five of these situations (indicated by asterisks in Table 2), such "low-fidelity" role playing sessions were used.

The key aspect of role playing is that it simulates interactions. It is not enough just to tell people about the roles and ask them to consider the interactions. When subjects were given information about the roles of the parties involved and were asked to consider this, it did not improve the accuracy of their forecasts (Armstrong, 1987).

Role playing would be relevant to trash-disposal fee problems. Various regulations could be presented to individuals who play the roles of household members. They would also be informed about the decisions of their neighbors. The government might respond to some of the consequences (e.g., illegal dumping or increased trash compacting by households) with new regulations, which, in turn, could be assessed. Town meetings could be simulated. Actions by trash collectors could also be predicted. The cost of such forecasts would be low compared with the cost of actually conducting and monitoring a trial of the various proposals.

Intention Surveys

Intention studies are surveys of individuals about what actions they plan to lake in a given situation or, if lacking a plan, what they expect to do. Such surveys are useful for predicting the outcomes of interventions. When a situation depends on the decisions of many people (such as with the trash collection for a community), surveys arc much more expensive than Delphi. However, they provide the perspective of those who will actually be making decisions. For example, consider the situation when the prohibition of Freon TM as a coolant was first proposed. Surveys might have been made of manufacturers of refrigerators and coolants to see how they would respond. In addition, one could have presented this situation to consumers and asked them how they would respond.

Tools for surveys have been improving since the 1936 *Literary Digest* poll predicted that Landon would easily defeat Roosevelt for president. Squire (1988), in a re-analysis of that event, concluded that the forecast was incorrect primarily because of nonresponse bias and secondarily because of sampling error. (People often assume that sampling error was the major cause.) Procedures for controlling sampling error are now well-known. Nonresponse error, where people fail to respond at all to the survey instrument, can be controlled by a variety of procedures, such as making extrapolations across waves (Armstrong and Overton, 1977*).* Perhaps the primary source of error is that caused by the nature of the response. Numerous improvements have been made to control for response error. Given these improvements, it is not surprising that the total error for political election forecasts decreased substantially in the United States from 1950 through 1978 (Perry, 1979).

Despite improvements in dealing with response bias, the problems for environmental forecasting are substantial. Citizens may have difficulty in predicting how an event or a change might affect them and in deciding how they will feel about the event. Lowenstein and Frederick (1997) discuss these issues and conclude that little evidence exists on the ability of people to predict how environmental changes will affect them. They did present evidence about how people would react to rain-forest destruction, restricted sport fishing because of pollution, and recovery of certain endangered species. They concluded that people greatly overestimate the effects of such changes on their life satisfaction.

As with other methods, objectivity is a key concern. When surveys are conducted by biased organizations. such as by political candidates, errors arc often substantial. Shamir (1986) classified 29 Israeli political surveys according to the independence of the pollster. The results showed that the more independent the pollster, the more accurate the survey.

When interventions would create large changes and where the behavior of decision makers is dependent upon decisions by others, respondents may find it difficult to predict how they would behave. Surveys are of less value in such cases.

Given all the ways that intentions or expectations may be wrong, it should not be surprising to find that sampling error alone provides a poor way to estimate prediction intervals. In a study of 56 trial polls in the 1992 presidential election, Lau (1994) concluded that the sample size of the poll was not closely related to the relative forecast errors for a set of surveys. When Buchannan (1986) examined errors for 155 political elections in nine countries from 1949 to 1985, they were twice as large as those expected from sampling error alone. This finding occurred with voting for political candidates, a behavior that was familiar to the respondents. For environmental concerns, where the future behavior may be less familiar to the respondents, one might expect that response and nonresponse biases would constitute large sources of error. These errors should be reflected in the assessment of prediction intervals.

Consider the trash-disposal fee problem again. If people have had no experience with such a system, it may be difficult for them to anticipate how they would behave. Their behavior depends to some extent on the behavior of their neighbors. Furthermore, if people did not want to comply with this new procedure and instead planned to illegally dis pose of trash, would they be willing to admit it in a survey? Could you imagine a respondent saying "Well, if it is going to cost that much to dispose of this waste legally, 1 will probably dispose of it illegally." Some procedures can help mitigate these problems. One way to do this is to use the random-response technique to help ensure confidentiality. For example, in a telephone survey, you could ask. "If a tax of 80 cents per bag of trash was implemented, would you dispose of any of your trash illegally? To answer, first flip a coin. If the coin turns up heads or if you expect that you would dispose of some trash illegally, then answer 'yes'." The amount of expected illegal disposal for the sample can then be teased out statistically. For example, if the average response for the respondents was 50 percent, then there would be no illegal dis posal expected; if it was 100 percent, then the assumption is that everyone would dispose of some trash illegally. Projective questions could also be used, such as, "Would your neighbors dispose of waste ille gally?" Sudman and Bradburn (1982, chap. 3) provide a discussion along with a 12-item checklist of how to ask threatening questions about behavior.

Delphi

Delphi involves the use of experts to make independent anonymous forecasts. Delphi goes beyond expert surveys in that it is conducted for two or more rounds. After the first round of forecasts, each expert receives a quantitative summary of the group's forecasts. In addition, anonymous explanations of their choices might be provided by the experts. Typically, two rounds are sufficient; however, if the cost associated with error is high, conducting three or four rounds may he worthwhile. Delphi is usually conducted by mail, and honoraria are paid to the participating experts. Stewart (1987) discusses the advantages and limitations of Delphi.

Delphi can be used to forecast trends, such as "What do experts expect to happen to the levels of New York City air pollution during the next 20 years?" It can also be used to forecast the effects of interventions: "What would be the impact of a $1/gallon federal tax on gasoline?" Experts need some level of domain expertise to make forecasts of change. Surprisingly, expertise beyond a modest level seems to have little relationship to accuracy (Armstrong, 1985, pp. 91-96). As a result there is little need to pay large honoraria to members of Delphi panels. Perhaps the primary criterion for the selection of experts for a Delphi panel is that they be unbiased.

Delphi requires only a few experts. The number of experts should be at least five but seldom more than 20 (Hogarth, 1978; Libby and Blashfield, 1978; Ashton and Ashton, 1985). As a result, Delphi studies can be rela tively inexpensive to conduct. This approach may be much less expensive than surveys that obtain information of individuals' intentions or expectations. For example, in predicting the outcomes of voter referendums on land use and property tax, Lemert (1986) found that, for the same level of accuracy, asking a few politicians for their predictions was more cost-effective than conducting a large-sample voter-intention survey.

When information is coming from a variety of sources, such as a number of Delphi respondents, the question comes up whether each source's information should be given the same weight. Rather than weighting by expertise, the preferred procedure is to weight each panel member's forecast equally, as long as each possesses at least some expertise. Based on studies to date, the required level of expertise is surprisingly low (Armstrong, 1985, pp. 91-96). Simple averages are commonly used and are often sufficient. McNees (1992) found little difference between the accuracy of means and medians in a study of economists' forecasts However, trimmed means (throwing out the highest and lowest estimates) are likely to be more accurate in cases involving high uncertainty. The median, the ultimate trimmed mean, may be the safest way to summarize forecasts (Larreche and Moinpour, 1983) if one has more than, say, 10 experts.

Delphi is relevant when data are lacking, the quality of the data are poor, or experts disagree with one another. As a result, Delphi is applicable when new interventions arc proposed or where a trend has recently undergone a shock. Nevertheless, judgments tend to be too conservative in the face of rapid change. In particular, judgment underestimates exponential growth (Wagenaar and Sagaria, 1975) and exponential growth is common in environmental problems. For example, Wagenaar and Timmons (1979) pre sented a computer-screen simulation of the growth of duckweed on a pond, an exponential process. Subjects asked to forecast when the pond would be covered greatly underestimated the time it would take. In another study (Wagenaar and Timmers, 1978), subjects were given information about pollution problems; when information was provided to subjects at more frequent time intervals, their predictions became less accurate. In a third study, Timmers and Wagenaar (1977) found that better judgmental predic tions were made when the variable reflected a decrease with time (e.g, instead of predicting population per square mile, predict square miles per person).

Because Delphi is based on (1) acting on prior research about the use of more than one expert, (2) using unbiased experts, (3) using structured questions, and (4) summarizing in an objective way, one would expect it to be more accurate than unaided judgment. It is. The few studies conducted on the validity of Delphi support its contribution to accuracy. Armstrong (1985, pp. 116-120); Stewart (1987); and Rowe et al. (1991) summarize these studies. Delphi is much more accurate than unaided judgmental fore casts. especially when the unaided forecasts arc made by only one or two people or where they are made in traditional group meetings.

Consider again the problem of trash collection. An impartial group of experts might be asked to predict what would happen if a fee were applied to trash containers. If the experts have direct experience with such systems in other localities or it they know the research literature on this topic, Delphi would seem to offer a reasonable way to forecast the effects of this policy.

One disadvantage of Delphi is that experts lend to be optimistic and overconfident: when they think about a problem, their confidence goes up much more rapidly than their accuracy. A tool that helps overcome this problem is the devil's advocate procedure, where someone is assigned for a short time to raise arguments about why the forecast or its interpretation might be wrong. The devil's advocate procedure led to more accurate forecasts in a study by Cosier (1978). Merely developing arguments against the validity of a forecast should produce a better assessment of confidence in a forecast (Koriat et al., 1980; Hoch 1985).

The variance among experts' forecasts offers a rough approximation of uncertainty (Ashton, 1985). For example, in McNees's (1992) examination of economic forecasts from 22 economists over 11 years, the actual values fell outside the range of their individual forecasts about 43 percent of the time. Little evidence exists on this topic, and it is not clear how to translate such information into a prediction interval.

For a more direct approach to an uncertainty estimate, one can ask each expert to provide 95 percent confidence intervals. However, experts are usually not well calibrated, and in some cases, about half of the estimates fall outside the 95 percent confidence intervals (Fischhoff and MacGregor, 1982; O'Connor and Lawrence. 1989). Experts are well calibrated when they receive good feedback about the accuracy of their forecasts. This issue is discussed by Plous (1993, chap. 19); he compares the excellent calibration for weather forecasters, who receive frequent, well-summarized feedback, to the poor calibration of physicians, who receive only occasional and poorly summarized feedback.

Analogies

To forecast the outcome of interventions, it is common for experts to search for cases where similar interventions have been conducted at different times or in different geographic areas and then to generalize from them. For example, some people generalize that socialist systems' poor environmental record is evidence that government regulation harms the environment. Such an assumption is counterintuitive to other people, who point out that socialist and free-market systems differ in many ways. The key point here is that the use of analogies is fraught with dangers.

Stewart and Leschine (1986) discuss analogies with respect to risk assessments. In making a decision about an oil refinery to be established at Eastport, Maine, the analysts rejected the use of worldwide estimates or tanker spills and instead relied on a comparison with one British port, Milford Haven. Although this decision-making group believed that this was a better comparison, one can reasonably attack the use of a single site as being risky because bias could easily enter into the selection of a single analogous case. To prevent such problems, it helps to select a large number of analogous situations. In the case of oil spills, it might be possible to rate all ports for similarity (without knowledge of their oil spill rates), then select a large sample of the most similar.

To picture how analogies might be properly used in an environmental decision process, consider the following problem. A community is considering alternative procedures for trash collection. Analogies might be useful if various trash-collection procedures had been tried in other communities and researchers had reported on the effects of these trials. Although each locality likes to think of itself as unique, a useful starting point would be to assume that people in a community would react to a given plan the same way others, on average, had reacted to similar plans in similar communities.

It is possible to structure the use of analogies by analyzing data from a sample of analogous situations. Fullerton and Kinnaman (1996) summarize some of these studies and report on the imposition of an $0.80 fee per 32-gallon can or bag of garbage in Charlottesville, Virginia. Their study examined the effects of this change on all key interest groups. The plan reduced the volume of garbage, but weight reductions were modest (only 14 percent). Illegal dumping also increased, so the true weight reduction was estimated at 10 percent. Considering administrative costs and the effects of illegal trash disposal, the program resulted in a net loss for the community. Such experience could guide a forecast of the effects of imposing such a fee in similar communities.

Conjoint Analysis

Conjoint analysts can be used to predict what strategy would be accepted. For example, one could propose different possible plans that would have various effects. The effects could be varied according to an experimental design. Once a model is developed, predictions can be made for changes in the design.

Judgmental Bootstrapping

Experts could be asked to predict the reactions to various possible interventions. A model could then he developed by regressing these predictions on the various elements of the intervention.

Extrapolation

Extrapolation involves making statistical projections using only the historical values for a time series; it is an appropriate tool to use when the causal factors will continue to operate as they have in the past. Furthermore, if one has little understanding of the causal factors, it might be best to use extrapolation.

Extrapolation has some useful characteristics. For one thing, it is fairly reliable. If agreement can be reached on the definition and length of the time series and on the statistical procedure, the same forecast will be achieved irrespective of who makes the forecast. Extrapolation can also be relatively simple and inexpensive. Although many complex procedures have been developed for extrapolation, such as the well-known Box-Jenkins methods, they have not produced gains in accuracy (Armstrong, 1984; Makridakis et al., 1993).

The opportunities for the introduction of biases in extrapolation are limited. Perhaps the major potential source of bias is that extrapolative forecasts can differ substantially depending on the time period examined. This bias can be reduced by selecting long time series and by comparing forecasts when different starting and ending points are used. Another source of bias associated with extrapolative forecasts involves the selection of the extrapolation method. To combat this bias, one should use simple, easily understood methods and preferably more than one method.

Extrapolation suffers when a time series is subjected to a shock or discontinuity. Few extrapolation methods account for discontinuities (Collopy and Armstrong, 1992b). Instead, when discontinuities occur, extrapolation will lead to large forecast errors. For example, nuclear power plant construction experienced a strong upward trend from 1960 through the mid-1970s and then a strong downward trend after that (Brown et al., 1994, p. 53). An extrapolation of nuclear power plant construction made in the early 1970s would have produced large errors.

There are many approaches to extrapolation. Most of them share the assumption that recent trends will continue. They vary primarily in how they weight the historical time periods. Exponential smoothing is widely used for this purpose. It is a moving average where the heaviest weight is placed on the most recent observation. Exponential smoothing is useful when one might expect a continuation of the forces that have operated in the past. It is less relevant for interventions, because these actions can change the direction or magnitude of the causal forces.

Exponential smoothing has several desirable qualities. First, it is simple and easy to understand. Second, it is inexpensive. And third, as noted, it puts more weight on the most recent data. However, this last benefit poses a limitation; that is, it is relevant only if the data have no seasonal effects (or have been seasonally adjusted) and if the most recent observations arc free of unusual events.

Different forms of exponential smoothing have been proposed, such as Brown's model to estimate the current smoothed average, $\overline{Y}_t$, using the latest value. Y_t, and the previous smoothed average $\overline{Y}_{t-1}$.

$$\overline{Y}_t = a\,Y_t + (1 - a)\ \overline{Y}_{t-1}$$

A smoothing factor (alpha) of 0.4 would put 40 percent of the weight for the level onto the latest period, Y_t, and 60 percent onto all preceding periods, $\overline{Y}_{t-1}$. This value means that 24 percent of the average (0.4 times 0.6) is applied to the period immediately preceding, with weights declining exponentially as older observations are treated. A similar procedure is used for estimating trends although the smoothing parameters will differ. (Monthly or quarterly data may first need to be adjusted for seasonal effects.) The need for seasonal adjustments is obvious in many cases, such as forecasts of electric-power demand. For a detailed discussion of exponential smoothing, see Gardner (1985).

Once the quantitative extrapolations have been made, it is risky to adjust the forecast judgmentally. Nevertheless, if those making the adjustments are unbiased and have good domain knowledge, and if the adjustments are made by a group of experts following structured procedures, then the adjustments are likely to improve accuracy. Even better is to use judgmental information as inputs to a quantitative model (Armstrong and Collopy, 1998).

Prediction intervals are easy to construct for exponential smoothing. The intervals should not be based on the fit to the data but, rather, on ex ante forecasts. Even so, these estimates are likely to underestimate uncertainty because they assume that the effects of causal factors will he the same in the future as they have been in the past.

Rule-Based Forecasting

When one has domain knowledge and large changes are involved, rule-based forecasting can be used. Rule-based forecasting is a validated, fully disclosed, and understandable set of conditional actions to make forecasts by assigning differential weights to extrapolation forecasts. In Collopy and Armstrong (1992a), domain knowledge and forecasting expertise led to a rule base with 99 rules conditioned on 18 features of time series. The features involved characteristics of the time series (such as the presence of a significant long-term trend), the amount of variability about the trend line, the presence of an unusual last observation, and so on. Some of these features are determined by judgment, although Adya et al. (1998) obtained good results by statistically determining all but four features: causal forces, irrelevant early data, cycles, and suspicious patterns. The rules assigned weights to four extrapolation methods to produce a combined forecast. These differential weights were shown to be more accurate than equal weights (Collopy and Armstrong 1992a; Adya of al., 1998). The key points arc that managers' knowledge should be applied to forecasting, and it should be done in a structured way.

Extrapolations typically ignore managers' domain knowledge. Rule-based forecasting integrates this knowledge by asking managers to describe their expectations about the future trend in a series. Those expectations represent the overall effects of the various causal forces that are acting on a series. To do this, managers would be asked to classify a series as growth, decay, supporting, opposing, regressing. or unknown. The forces are listed in Table 3 along with some examples, and the procedure is described in Armstrong and Collopy (1993). For example, a manager's expectation that automobiles will produce less pollution per gallon of fuel (a decay series) should be reflected in the forecast. Causal forces provide a simple and inexpensive way to use domain expertise when making statistical extrapolations.

Table 3. Relationship of causal forces to trends.

Type of causal forces	Causal forces direction when … Trend has been up	Trend has been down	Examples
Growth	Up	Up	Gross national product; electricity consumption
Decay	Down	Down	Resource prices
Supporting	Up	Down	Short-term land prices?
Opposing	Down	Up	Wildlife
Regressing	(Toward a mean)		Demographic (% male births)

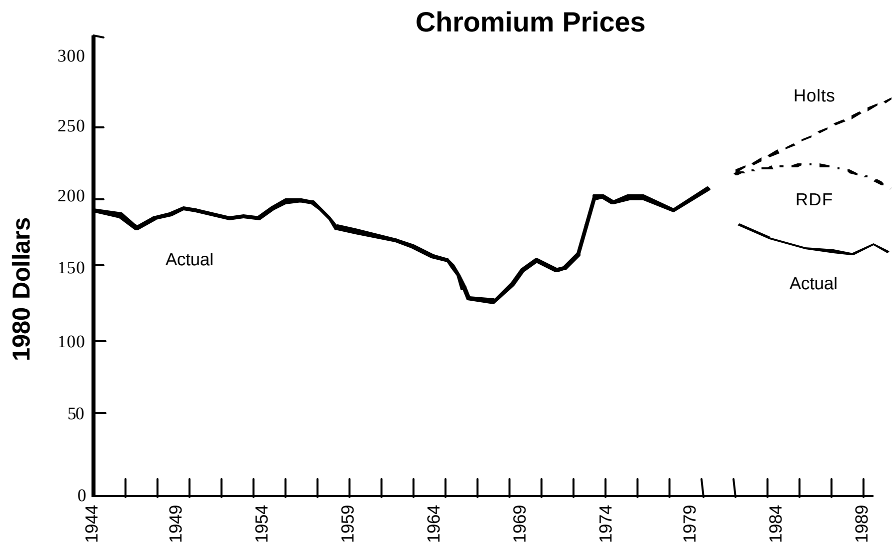
One rule for the use of causal forces is to avoid extrapolating a trend if it would be contrary to the expected trend. The expected trend is based on a specification of the causal forces affecting the series. Consider the situation in 1980, when Julian Simon made the following challenge: "Pick any natural resource and any future date. I'll bet the [real] price will not rise." He based this on long-term trends and argued that there had been no major changes in the long-term causal factors. Paul Ehrlich, an ecologist from Stanford University, accepted the challenge; he selected 10 years and five metals (copper, chromium, nickel, tin, and tungsten) whose prices had been rising in recent years. However, the causal forces for the prices of resources are "decay" because of improved procedures for prospecting, more efficient extraction procedures, lower energy costs, reduced transportation costs, development of substitutes, more efficient recycling methods, and more open trade among nations. The exhaustion of resources might lead to increased prices; however, this seldom has a strong effect because new sources are found. For example, Ascher (1978, pp. 139-141) showed that forecasts of the ultimate available petroleum reserves *increased* from the late 1940s to the mid-1970s. Such changes seem common for resources because of improvements in exploration technology. Thus, in my judgment, the overall long-term causal force is decay, so metals prices would be expected to decrease. They did, and Simon won the bet; his predictions were correct for all live metals (Tierney, 1990).

Rule-based forecasting is especially useful when domain knowledge indicates that recent trends may not persist. In the case of metals forecasting, Ehrlich assumed that recent price trends would continue. I implemented this assumption in Figure 2 by using Holt's exponential smoothing to extrapolate recent trends for one of his five metals, chromium, and obtained a forecast of sharply rising prices.[1] In contrast, although the rule base initially forecasts an increase in prices (because it allows that short-term trends might continue), over the 10-year horizon, the forecast becomes dominated by the long-term trend, which is downward and consistent with the causal forces (see Figure 2, next page). This same pattern was found for each of the five metals for forecasts made in 1980.

Much work is currently being done on the integration of judgment and statistical forecasting. Accuracy is almost always improved if the integration uses unbiased and structured inputs and if the judgmental inputs are made independently of the statistical forecasts (Armstrong and Collopy, 1998).

[1] The forecasts were prepared by Monica Adya, using a version of rule-based forecasting that is described in Adya, Armstrong, Collopy, and Kennedy (1998). The data were obtained from *Metals Week*.

Figure 2. Actual commodity prices for chromium from 1944 to 1990 and prices forecasted for the period 1981 to 1990 by extending the trend with Holt's exponential-smoothing technique and by rule-based forecasting.



Expert Systems

Expert systems seem ideal for cases involving environmental forecasting. One can draw upon the expertise of the best experts. If econometric models have been developed, such as the above-mentioned Turner et al. (1992) study, the resulting information about relationships could be incorporated into the expert system. Further refinements could he made by quantifying experts' rules by judgmental bootstrapping. Information about citizen responses could be incorporated by using conjoint studies. Thus, expert systems allow for the systematic and explicit integration of all extant knowledge about a situation. Expert systems are being used for a variety of problems. Unfortunately, information about the predictive validity of expert systems is limited. but positive.

Econometric Models

Econometric models use information about causal relationships to make forecasts. The causal relationships should be specified by using domain knowledge (i.e., information that a manager has about the problem). Well-established theories should also be used: thus, we know that income should, in most cases, be positively related to demand for an item, and price should he negatively related. Given a description of the product and market, we can also use prior research to determine the approximate magnitude of the relationship. So if a community makes it more expensive to pollute, one would expect less pollution if the plan is properly designed, and perhaps more graft if the plan is poorly designed. Theory or domain knowledge can he used to identify key variables, specify the direction and form of the relationships, and set limits on the values that coefficients may take.

The value of well-established theories should normally take precedence over domain knowledge. For example, Winston (1993) examined 30 published studies where economists, using theory, made predictions about the effects of deregulation. The predictions in these studies conflicted with the opinions of the people affected by the deregulation. The economists predicted that deregulation would, in general, be good for consumers, whereas those who would be affected by the change predicted the opposite. As it turned out, the economists' predictions were almost always correct.

While extrapolations assume that everything continues as in the past, econometric models assume only that the relationships will remain constant. Given an estimate of the relationship of the causal variables to the dependent variable, one must forecast changes in the causal variables in order to calculate a forecast for the variable of interest.

For example, to forecast changes in the level of automobile pollution, one might need to forecast the number of miles driven, average vehicle weight, average speed, engine efficiency, fuel type, and effectiveness of emissions-control equipment. If a causal variable changes direction or if it changes at a much different rate than it has in the past, the econometric model will reflect this in its forecasts. An econometric model can also estimate the effects of potential changes such as a new type of engine, a new regulation on automobiles, or a large change in the tax on gasoline.

Methods that do not use domain knowledge or theory, such as step-wine regression, should not be used for forecasting (Armstrong. 1985, pp. 223-225). Not only might the analysts ignore useful information, but also they may be misled by spurious correlations. Also, I see little hope for neural nets, despite their current popularity. Chatfield (1995) discusses the limitations of neural nets.

Econometric methods arc most useful where (1) large changes are expected; (2) a priori information about relationships is strong; (3) good data are available; and (4) causal factors arc easier to forecast than the variable of interest. These conditions are often encountered in forecasting. For example, over the long-term, the effects of air pollution are likely to be substantial. Studies about the causes of pollution provide a reasonable level of knowledge about the causal relationships. Some of the causal variables, such as population and production of various goods, can be forecast more easily than one could directly forecast air pollution. In addition, relationships can sometimes be estimated by laboratory or field experiments. as illustrated by Turner et al. (1992). Finally, the quality of the data is improving. Not surprisingly, then, many researchers use econometric methods to forecast various types of environmental impacts.

One of the major advantages of econometric methods, in comparison with other forecasting methods, is that alternate interventions can be compared with one another. In effect, one is comparing results in an objective way with an attempt to hold all other influences constant. Turner et al. (1992) used this approach in their 50-year forecasts of acid rain to examine the effects that different policies might have.

The technology for econometric methods has become much more complicated since the least-absolute-value method was introduced in 1757, followed by the least-squares method in 1805. But highly complex procedures are not easily understood by decision makers. Worse, little validation research has been conducted on complex procedures. What has been done suggests that complexity seldom leads to improved accuracy. Dielman (1986), using simulated data, concluded that the least-absolute-value method still works well, especially for data that suffer from outliers. In general, theory-based single-equation ordinary least-squares methods have forecast well when compared with alternative procedures.

Simple econometric models aid understanding andd reduce the potential for errors. The benefits seem to translate into practice: in a field study involving the forecasting of state-government revenues, the use of simple econometric models was associated with improvements in accuracy, while the use of complex econometric models was associated with reduced accuracy (Bretschneidcr cl al., 1959).

Mechanical adjustments are often necessary to adjust for errors in the current status (i.e., to adjust the starting value). For example, one useful procedure is to add half of the latest error to the forecast. Once an econometric forecast has been prepared, generally speaking, it should not be adjusted judgmentally (Armstrong and Collopy 1998). However, if there has been a major recent event that has not yet been reflected in the data, structured judgmental procedures might be used to adjust the level.

Econometric methods can aid in the construction of confidence intervals. Such intervals are expected to be underestimated if, as is almost always the case, they make no provision for the uncertainty involved with predicting causal variables or for the possibility that relationships might change. Thus, the use of the traditional standard error of a model as the foundation for estimating prediction intervals should he supplemented by other approaches. This practice is illustrated by Turner el al. (1992), who compared forecasts from different models and also compared forecasts given different assumptions about the forecasts of the causal variables. In addition, they examined limitations of the models, such as the effects of excluded variables. Excluded variables seemed to be a serious limitation, although their effects were not quantified. They then tested their model in different geographic regions. Finally. they tested the model by making long-term backcasts to prehistoric times and compared the results with independently obtained lake-chemistry estimates.

If many communities have tried different plans for trash disposal, an econometric model might be estimated to predict the outcomes of various plans. The econometric model could help to control for differences among communities and also for factors that change with time. Such a model would aid in determining the effects of alternative trash fees.

Selecting and Combining Forecasts

Assuming that data exist for using each of the above forecasting methods. which method should be used? If factors that caused changes in the past continue to operate in the same way in the future, the choice of a method is not so important; each method would be expected to have reasonably good accuracy. But given the large changes expected in many environmental problems, the selection of a forecasting method is likely to be important.

Judgment is helpful for estimating levels, while extrapolation and econometric models are better at forecasting changes. Extrapolation is good at forecasting changes when the causal factors continue to operate as in the past, whereas econometric models can compensate for substantial changes in the causal forces. As a result, econometric forecasts are generally more accurate than extrapolation or judgment when large changes are involved (Armstrong, 1985, pp. 391-420). Fildes's (1985) review adds further support and also suggests that econometric models provide small improvements in accuracy for short-range forecasts. Note that these studies were, for the most part, conducted in situations that did not involve environmental forecasting. However, Ascher (1978, p. 119). in examining 10-year forecasts of electricity consumption, concluded that extrapolation and econometric methods were more accurate than judgment. Also, Rausser and Oliveira (1976), in a study of wilderness-area use, found that econometric methods were more accurate than extrapolations and that a combination of econometric forecasts was even more accurate. In general, assuming adequate data and a good understanding of causal relationships, econometric methods would be the preferred forecasting method because they use much relevant information in a structured way.

Rather than trying to chose the single best method, the problem is better framed by asking which methods would help to improve accuracy. Baker et al. (1980) illustrated this use of multiple forecasting methods. They forecasted the impact that offshore nuclear power plats would have on beach visitation by using expert surveys, analogies (visits to beaches near land-based nuclear plants), and intentions of potential beach visitors.

Combined forecasts are those where one uses different methods to make forecasts for the same situation and then combines the forecasts. Combined forecasts are especially useful where much uncertainty exists about which method is likely to produce the most accurate forecasts. Combined forecasts typically improve accuracy because each forecasting method makes some contribution.

Much research suggests that combined forecasts are generally more accurate than forecasts prepared with a single method. Furthermore. they are sometimes more accurate than the best component. Combining also offers protection against mistakes because their effects are muted by the other forecasts. Finally, combining forecasts from different methods and data will add to objectivity and to the appearance of objectivity.

Combining of forecasts should be done mechanically to help assure users that the procedure is objective. That is, a rule should be used, and it should be fully described. An example would be "equal weights," which states that one adds each of the forecasts and calculates an average. This objectivity in the weighting process is expected to improve accuracy, and equal weights is robust across situations (Clemen 1989). For example, Bretschneider et al. (1989), in a field study, found that U.S. states that used mechanical combinations of forecasts had more accurate revenue forecasts than those using subjective combinations.

Uncertainty

The assessment of uncertainty in forecasting should not include tests of statistical significance because they do not relate well to issues of importance in forecasting and because they are so often misinterpreted (McCloskey and Ziliak, 1990). Instead, one can provide prediction intervals. The prediction interval represents the proportion of times that tihe actual forecasts are expected to fall within a specified range. Thus, 95 percent prediction intervals should be expected to contain 95 percent of the true values.

Estimates of prediction intervals might be obtained by comparing forecasts from different methods. While agreement inspires more confidence and disagreement less, the translation of these differences to prediction intervals must be done subjectively. Some improvements might be achieved if the prediction intervals are estimated independently by a number of approaches, and the estimates are then combined mechanically.

To develop prediction intervals, it is generally best to make forecasts by simulating the situation facing the forecaster and then calculating ex ante forecast errors that can be used to construct prediction intervals for each forecast horizon. The resulting limits can be smoothed over the forecast horizon.

Using the Forecasts

Interestingly, researchers, educators, forecasters, and decision makers all use similar criteria for judging which forecasting models are most useful (Yokum and Armstrong, 1995). Accuracy is generally rated as the most important criterion. These experts also agree that ease of understanding and ease of use are nearly as important as accuracy. These agreements on criteria suggest that it might be possible to reach agreement of what *forecasting methods* perform best for a given situation.

Forecasts are used by decision makers, politicians, special-interest groups, manufacturers, lawyers, and the media. Given their different needs, they may desire different forecasts. So agreement on *forecasts* is a difficult matter, especially if no prior agreement has been reached with the decision makers about the proper forecasting methods and if the forecast is surprising to some.

Unfortunately, adjustments to forecasts are often made by biased experts. In a survey of members of the International Institute of Forecasters, respondents (n = 269) were given the following statement: “Too often, company forecasts are modified because of political considerations.” On a scale from 1 = “disagree strongly” to 7 = “agree strongly,” the average response was 5.37.[2] Fildes and Hastings (1994). in an intensive study of forecasting in a large multidivision firm, found 64 percent of their respondents agreeing that forecasts are frequently modified for political reasons.

Subjective adjustments may expose the forecaster to charges of bias. Glantz (1982) describes how a subjective adjustment of a weather forecast led to the prediction of an extreme drought in Yakima in 1977. Farmers took appropriate action for such a drought, but, as it turned out, the drought did not occur. Serious losses resulted from the farmers’ actions. As a result, the farmers sued the government, and the subjective adjustment was challenged as evidence of malpractice.

Often, the most useful forecasts tell us something new, that is, they challenge existing expectations. They arc valuable because they allow us to take corrective action. However. such forecasts are frequently ignored or resisted by organizations, as found in Griffith and Wellman’s (1979) study of the need for hospital beds. Fortunately, there are tools to aid users of forecasts in such situations. These tools involve presentation techniques and scenarios.

Presentation Techniques

If decision makers are biased to favor certain forecasts, it would seem natural for them to suspect that the forecasting tools are inadequate when forecasts are not to their liking. Thus, in presentations to decision makers, do

[2] This survey was conducted in 1989 by Thomas Yokum and me. The responses to this question were analyzed depending on whether the respondents identified himself or herself as primarily a decision maker, practitioner, educator, or researcher. While the practitioners stated the strongest agreement, there were no statistically significant differences among these groups.

not start off with the forecasts. The initial emphasis should be on the forecasting methods in order to gain agreement on which methods are appropriate.

Forecasts involve uncertainty, so prediction intervals should be presented along with forecasts. Planners can then prepare contingencies depending on the range of possible futures. Unfortunately, it is common to report only expected values and not prediction intervals. Rush and Page (1979) examined 372 long-term metals forecasts from 1910 to 1964. Explicit re ferences to uncertainly were not the rule. Furthermore, their use decreased from 22 percent of the forecasts published up to 1939 to only eight percent afterward. In a survey of "marketing/forecasting managers" at 134 U.S. firms (Dalrymple, 1987), fewer than 10 percent said that they usually used prediction intervals, and almost half said that they never used them.

Although the capability exists to provide entertaining graphics along with forecasts, these may not improve the message. Wagenaar et al. (1985) compared the delivery of weather forecasts by radio and TV. They concluded that recall was not improved by TV except when written summary statements were also provided on TV.

People tend to understand and remember examples. Thus, it makes sense to reinforce your forecasts with vivid examples. In a study about firefighters' preference for risk, Anderson (1983) found that the use of concrete examples was more effective than was the presentation of relevant statistical results. An experiment by Read (1983) suggests that politicians may be more influenced by a single historical event (e.g., "Here is what happened at Three-Mile Island") than by a generalization from a broad range of situations.

Scenarios

Forecasts that call for changes are often resisted by decision makers. For example, Baker (1979) found that hurricane warnings are frequently ignored. People do not like to receive information about the potential destruction of their homes, and they tend to revise the forecast to make it less threatening.

The use of scenarios can help decision makers deal with forecasts that have unpleasant consequences. A scenario is a story about what happened in the future. The choice of tenses in the preceding sentence was intentional. The use of the past tense helps to add realism and to gain the decision maker's commitment to a course of action for a given forecast.

Scenarios are likely to lead people to overestimate the likelihood of an event. Certainly, the forecaster should point out that the function of the scenario is not to forecast but to decide how to use forecasts. The fact that the event will seem more likely should aid people to take unfavorable forecasts more seriously.

Research by psychologists suggests that effective scenarios:

- Use concrete examples
- Include events that are representative of what one might expect
- Link the events by showing causal relationships
- Ask the decision makers to project themselves into the situation
- Ask the decision makers to describe how they acted in this scenario.

Additional details on the scenario procedure arc provided in Armstrong (1985, pp. 40-45).

Scenarios allow decision makers to report (predict) how they would act given certain future circumstances. People's responses about their intentions may affect their subsequent behavior (Greenwald et al., 1987). Such predictions of behavior have been shown to affect people's behavior when similar situations have been encountered later (Grcgory et al., 1992). The generation of scenarios might also be beneficial because people tend to predict that they will act in socially responsible and rational ways. Thus, by asking citizens to describe how they would react to a new trash disposal plan, one might affect the respondents' behavior in such a way as to increase the plan's chances for success if they liked the plan.

Baker, E. J. (1979), “Predicting responses to hurricane warnings: A reanalysis of data from four studies,” *Mass Emergencies*, 4, 9-24.

Baker, E. J. et al. (1980), “Impact of offshore nuclear power plants: Forecasting visits to nearby beaches,” E*nvironment and Behavior,* 12, 367-407.

Blattberg, B. C. and Hoch, S. J. (1990), “Database models and managerial intuition: 50 percent model + 50 percent manager,” *Management Science,* 36, 887-899.

Brenner, L. A., Koehler, D. J., and Tversky, A. (1996), “On the evaluation of one-sided evidence,” *Journal* of *Behavioral Decision Making,* 9, 59-70.

Bretschneider, S. I., et al. (1989), “Political and organizational influences on the accuracy of forecasting state government revenues,” *International Journal of Forecasting,* 5, 307-319.

Brown, L. R., Kane, H., and Roodman, D. M. (1994), *Vital Signs 1994*. New York: Norton.

Buchannan, W. (1986), “Election predictions: An empirical assessment,” *Public Opinion Quarterly* 50, 222-227.

Cerf, C. and Navasky. V. (1984), *The Experts Speak*. New York: Pantheon.

Chatfield, C. (1995), “Positive or negative?” *International Journal of Forecasting*, 11, 501-502.

Clemen, R. (1989), “Combining forecasts: A review and annotated bibliography,” *International Journal of Forecasting*, 5, 559-583.

Collopy, F. and Armstrong, J. S. (1992a),” Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations,” *Management Science,* 38, 1394-1414.

Collopy, F. and Armstrong, J. S. (1992b), “Expert opinions about extrapolation and the mystery of the overlooked discontinuities.” *International Journal of Forecasting,* 8, 575-582.

Cosier, R. A. (1978), “The effects of three potential aids for making strategic decisions on prediction accuracy,” *Organizational Behavior and Human Performance*. 22, 295-306.

Dalrymple, D. J. (1997), “Sales forecasting practices: Results from a United States survey,” *International Journal of Forecasting*, 3, 379-391.

Dielman, T. E. (1986), “A comparison of forecasts from least absolute value and least squares regression,” *Journal of Forecasting,* 5, 189-195.

Fildes, R. (1985), “The state of the art: Econometric models,” *Journal of the Operational Research Society*, 36, 549-586.

Fildes, R. and Hastings, R. (1994), “The organization and improvement of market forecasting,” o *Journal of the Operational Research Society*, 45, 1-16.

Fischhoff, B. and MacGregor, D. (1982), “Subjective confidence in forecasts,” *Journal of Forecasting,* 1, 155-172.

Fullerton, D. and Kinnaman, T. C. (1996), “Household responses to pricing garbage by the bag,” *American Economic Review,* 86, 971-984.

Gardner, E. S. (1984), “The strange case of the lagging forecasts.” *Interfaces,* 14, 47-50.

Weimann, G. (1990), "The obsession to forecast: Pre-election polls in the Israeli press," *Public Opinion Quarterly*, 54, 396-408.

Winston, C. (1993), "Economic deregulation: Days of reckoning for microeconomists," *Journal of Economic Literature,* 91, 1263-1289.

Yokum, J. T. and Armstrong, J. S. (1995), "Beyond accuracy: Comparison of criteria used to select forecasting methods," *International Journal of Forecasting,* 11, 591-597.