



University of Pennsylvania
ScholarlyCommons

Marketing Papers

Wharton School

2001

Selecting Forecasting Methods

J. Scott Armstrong

University of Pennsylvania, armstrong@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

 Part of the [Marketing Commons](#)

Recommended Citation

Armstrong, J. S. (2001). Selecting Forecasting Methods. Retrieved from http://repository.upenn.edu/marketing_papers/147

Suggested Citation:

Armstrong, J.S. Selecting Forecasting Methods. In *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Ed. J. Scott Armstrong). Kluwer, 2001.

Publisher URL: <http://www.springer.com/business+%26+management/business+for+professionals/book/978-0-7923-7930-0>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/marketing_papers/147

For more information, please contact libraryrepository@pobox.upenn.edu.

Selecting Forecasting Methods

Abstract

I examined six ways of selecting forecasting methods: Convenience, “what’s easy,” is inexpensive, but risky. Market popularity, “what others do,” sounds appealing but is unlikely to be of value because popularity and success may not be related and because it overlooks some methods. Structured judgment, “what experts advise,” which is to rate methods against prespecified criteria, is promising. Statistical criteria, “what should work,” are widely used and valuable, but risky if applied narrowly. Relative track records, “what has worked in this situation,” are expensive because they depend on conducting evaluation studies. Guidelines from prior research, “what works in this type of situation,” relies on published research and offers a low-cost, effective approach to selection. Using a systematic review of prior research, I developed a flow chart to guide forecasters in selecting among ten forecasting methods. Some key findings: Given enough data, quantitative methods are more accurate than judgmental methods. When large changes are expected, causal methods are more accurate than naive methods. Simple methods are preferable to complex methods; they are easier to understand, less expensive, and seldom less accurate. To select a judgmental method, determine whether there are large changes, frequent forecasts, conflicts among decision makers, and policy considerations. To select a quantitative method, consider the level of knowledge about relationships, the amount of change involved, the type of data, the need for policy analysis, and the extent of domain knowledge. When selection is difficult, combine forecasts from different methods.

Disciplines

Business | Marketing

Comments

Suggested Citation:

Armstrong, J.S. Selecting Forecasting Methods. In *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Ed. J. Scott Armstrong). Kluwer, 2001.

Publisher URL: <http://www.springer.com/business+%26+management/business+for+professionals/book/978-0-7923-7930-0>

Selecting Forecasting Methods

J. Scott Armstrong

The Wharton School, University of Pennsylvania

ABSTRACT

I examined six ways of selecting forecasting methods: **Convenience**, “what’s easy,” is inexpensive, but risky. **Market popularity**, “what others do,” sounds appealing but is unlikely to be of value because popularity and success may not be related and because it overlooks some methods. **Structured judgment**, “what experts advise,” which is to rate methods against prespecified criteria, is promising. **Statistical criteria**, “what should work,” are widely used and valuable, but risky if applied narrowly. **Relative track records**, “what has worked in this situation,” are expensive because they depend on conducting evaluation studies. **Guidelines from prior research**, “what works in this type of situation,” relies on published research and offers a low-cost, effective approach to selection. Using a systematic review of prior research, I developed a flow chart to guide forecasters in selecting among ten forecasting methods. Some key findings: Given enough data, quantitative methods are more accurate than judgmental methods. When large changes are expected, causal methods are more accurate than naive methods. Simple methods are preferable to complex methods; they are easier to understand, less expensive, and seldom less accurate. To select a judgmental method, determine whether there are large changes, frequent forecasts, conflicts among decision makers, and policy considerations. To select a quantitative method, consider the level of knowledge about relationships, the amount of change involved, the type of data, the need for policy analysis, and the extent of domain knowledge. When selection is difficult, combine forecasts from different methods.

KEYWORDS: Accuracy, analogies, combined forecasts, conjoint analysis, cross-sectional data, econometric methods, experiments, expert systems, extrapolation, intentions, judgmental bootstrapping, policy analysis, role playing, rule-based forecasting, structured judgment, track records, and time-series data.

How should one select the best method for producing a forecast? Chambers, Mullick and Smith (1971) provided answers with a fold-out chart. The chart, based on their opinions, had six descriptors down the first column that were a mix of objectives and conditions (e.g., accuracy, applications, data required, and cost of forecasting). Across the top, it had 18 forecasting techniques, some of which overlapped with others (e.g., regression, econometric methods). During the following 17 years, the *Harvard Business Review* sold over 210,000 reprints of the article, making it one of its most popular reprints. Chambers, Mullick and Smith (1974) expanded upon the article in a book. Since then, much has been learned about selecting methods.

I examine six ways to select forecasting methods: convenience, market popularity, structured judgment, statistical criteria, relative track records, and guidelines from prior research. These approaches can be used alone or in combination. They may lead to the selection of more than one method for a given situation, in which case you should consider combining forecasts.

CONVENIENCE

In many situations, it is not worth spending a lot of time to select a forecasting method. Sometimes little change is expected, so different methods will yield similar forecasts. Or perhaps the economics of the situation indicate that forecast errors are of little consequence. These situations are common.

Convenience calls to mind the Law of the Hammer (give a child a hammer and he will find many things that need to be pounded). There is a common presumption that researchers who are skilled at a technique will force their technique on the problem at hand. Although this has not been studied by forecasters, related research by psychologists, on selective perception, supports this viewpoint.

Convenience may lead to methods that are hard to understand. Statisticians, for example, sometimes use Box-Jenkins procedures to forecast because they have been trained in their use, although decision makers may be mystified. Also, a method selected by convenience may lead to large errors in situations that involve large changes.

MARKET POPULARITY

Market popularity involves determining what methods are used by other people or organizations. The assumptions are that (1) over time, people figure out what methods work best, and (2) what is best for others will be best for you. Surveys of usage offer only indirect evidence of success.

Dalrymple (1987), using a mail survey, obtained information about the usage of forecasting methods at 134 companies in the U.S. Exhibit 1 shows information from his study. He also cited other studies on the usage of sales-forecasting methods and these contained similar findings.

Exhibit 1
Sales Forecasting Methods Used by Firms

	Regularly Used as Percentage		Regularly Used as Percentage
Expert Opinion		Extrapolation	
<i>Internal</i>		Naive	30.6
Sales force	44.8	Moving average	20.9
Executives	37.3	Rate of change (percentage)	19.4
<i>External</i>		Rate of change (units)	15.7
Industry survey	14.9	Exponential smoothing	11.2
Analogies		Regression against time	6.0
Leading indicators	18.7	Box-Jenkins	3.7
Econometric			
Multiple regression	12.7		
Econometric methods	11.9		

Additional studies have been conducted since Dalrymple's. Sanders and Manrodt (1994), for example, found that while knowledge of quantitative methods seemed to be improving, firms still relied heavily on judgmental methods.

Frank and McCollough (1992) surveyed 290 practitioners from the Finance Officers Association (for U.S. state governments) in 1990. The most widely used forecasting method was judgment (82% of the respondents), followed by trend lines (52%), regression (26%), moving averages (26%), and exponential smoothing (10%).

Rhyne (1989) examined forecasting practices at 40 hospitals by interviewing senior management. Judgmental methods were commonly used: 87% reported using the 'jury of executive opinion' and 67.5% used expert forecasts. Given the political nature of hospital forecasts, their use of judgmental methods would seem to present serious problems with bias. For quantitative methods, 52.5% of the hospitals used moving averages, 12.5% used exponential smoothing, and 35% used regression.

One of the problems with usage surveys is that forecasting techniques have not been clearly defined. For example, what does "simple regression" mean? It might mean regression against time, but not all respondents would use this definition.

Another problem is that the conditions are not always described. This is difficult to do, and in fact, researchers have rarely even requested such information. Dalrymple (1987) and Mentzer and Cox (1984) are among the few who did. They examined methods that firms used for short-, medium-, and long-term sales forecasting (e.g., their respondents seldom used extrapolation for long-range forecasts), those used for industrial goods versus consumers goods (e.g., industrial firms placed more reliance on sales-force opinions), and those used by small or large firms (large firms used more quantitative methods). However, even these distinctions are too broad to be of much use. Forecasters need to know specifics about the forecasting task, such as the methods firms use to forecast new-product sales for consumer durables when at the concept phase, or how should one forecast competitor’s actions?

Another limitation of usage studies is that they have not measured success. They measure only usage (actually, they measure only perceived usage reported by people who would like to be regarded as good managers). If firms do not conduct evaluations of alternative methods (and few do), usage offers a poor guide to what *should* be done. Sometimes firms assume that methods are effective and use them widely even when they are of no value. Certainly, usage is unrelated to efficacy in many cases. Forecasters use expert opinions even when ample evidence exists that other methods are more accurate, as described for judgmental bootstrapping (Armstrong 2001b) and econometric methods (Grove and Meehl 1996).

In some cases, what is done does not agree with experts’ belief about what *should* be done. Armstrong, Brodie and McIntyre (1987) surveyed forecasting practitioners, marketing experts, and forecasting experts concerning how to forecast competitors’ actions. What practitioners did differed from what marketing experts recommended, which, in turn, differed from what forecasting experts preferred (Exhibit 2). For example, practitioners seldom used game theory although almost half of the marketing experts thought it useful (few forecasting experts agreed). Similarly, the use of role playing was minimal although it was one of the forecasting experts’ highest-rated methods in this situation.

Exhibit 2
Usage Can be a Poor Guide to Selection of Forecasting Methods:
Percentages Using or Preferring Methods to Forecast Competitors’ Actions

Methods to Forecast Competitors’ Actions	% Usage	% Experts’ preferences	
	Practitioners (n=59)	Marketing (n=15)	Forecasting (n=18)
Expert opinion (experts who know about the situation)	85	100	83
Extrapolation (statistical analysis of analogous situations)	58	53	50
Intentions (ask competitors)	22	60	33
Experimentation (try the strategy on a small scale)	17	60	22
Game theory (formal use of game theory)	8	47	22
Role-playing (formal acting out of the interactions involved)	7	20	61

Finally, surveys have typically overlooked methods such as role playing, judgmental bootstrapping, conjoint analysis, and expert systems. As a result, market popularity is the enemy of innovation.

STRUCTURED JUDGMENT

When a number of criteria are relevant and a number of methods are possible, structured judgment can help the forecaster to select the best methods. In structured judgment, the forecaster first develops explicit criteria and then rates various methods against them.

Evidence that structured judgments are superior to unstructured judgments has been found for many types of selection problems. For example, in a review of research on the selection of job candidates, Campion, Palmer and

Campion (1997) concluded “In the 80-year history of published research on employment interviewing, few conclusions have been more widely supported than the idea that structuring the interview enhances reliability and validity.”

☛ **List the important criteria before evaluating methods.**

Yokum and Armstrong (1995) summarized selection criteria that had been examined in earlier surveys by Carbone and Armstrong (1982), Mahmoud, Rice and Malhotra (1986), Mentzer and Cox (1984), and Sanders and Mandrodt (1994). They also reported findings from an expert survey of 94 researchers, 55 educators, 133 practitioners, and 40 decision makers. The results (Exhibit 3) were similar to those from the previous studies. The earlier studies did not include the ability of the forecasting model to compare alternative policies, the ability to make forecasts for alternative environments, and learning. Learning means that, as forecasters gain experience, they improve their forecasting procedures.

Exhibit 3
Criteria for Selecting a Forecasting Technique
(scale: 1 unimportant to 7 unimportant)

Criteria	Mean Importance Rating (number responding)				
	Researcher (94)	Educator (55)	Practitioner (133)	Decision Maker (40)	Average (322)
Accuracy	6.39	6.09	6.10	6.20	6.20
Timeliness in providing forecasts	5.87	5.82	5.92	5.97	5.89
Cost savings resulting from improved decisions	5.89	5.66	5.62	5.97	5.75
Ease of interpretation	5.54	5.89	5.67	5.82	5.69
Flexibility	5.54	5.35	5.63	5.85	5.58
Ease in using available data	5.59	5.52	5.44	5.79	5.54
Ease of use	5.47	5.77	5.39	5.84	5.54
Ease of implementation	5.24	5.55	5.36	5.80	5.41
Incorporating judgmental input	4.98	5.12	5.19	5.15	5.11
Reliability of confidence intervals	5.09	4.70	4.81	5.05	4.90
Development cost (computer, human resources)	4.70	5.02	4.83	5.10	4.86
Maintenance cost (data storage, modifications)	4.71	4.75	4.73	4.72	4.73
Theoretical relevance	4.81	4.20	4.43	3.72	4.40
Ability to compare alternative policies	—	—	—	—	—
Ability to examine alternative environments	—	—	—	—	—
Learning	—	—	—	—	—

Decision makers, practitioners, educators, and researchers had similar views on the importance of various criteria as seen in Exhibit 3. The average rank correlation among these groups is .9.

All the surveys showed that accuracy is the most important criterion. Mentzer and Kahn (1995), in a survey of 207 forecasting executives, found that accuracy was rated important by 92% of the respondents. However, the relative importance of the various criteria depends upon the situation. The importance ratings varied for short versus long series, whether many or few forecasts were needed, and whether econometric or extrapolation methods were involved. For example, for forecasts involving policy interventions, the experts in Yokum and Armstrong’s (1995) survey rated cost savings from improved decisions as the most important criterion.

☛ **Assess the method’s acceptability and understandability to users.**

Although most academic studies focus on accuracy, findings from previous surveys indicate that ease of interpretation and ease of use are considered to be nearly as important as accuracy (see Exhibit 3). It does little good to propose an accurate method that will be rejected or misused by people in an organization. Confidential surveys of users can help to assess the acceptability and understandability of various methods.

↳ **Ask unbiased experts to rate potential methods.**

To find the most appropriate methods, you should ask a number of experts to rate various forecasting methods. The experts should have good knowledge of the forecasting methods and should have no reason to be biased in favor of any method. The experts also should be familiar with the specific forecasting situation. If outside experts are used, they should be given written descriptions of the situation. This would aid them in making their evaluations and will provide a historical record for future evaluations. Formal ratings should be obtained independently from each expert. The Delphi procedure (Rowe and Wright 2001) provides a useful way of obtaining such ratings.

STATISTICAL CRITERIA

Statisticians rely heavily upon whether a method meets statistical criteria, such as the distribution of errors, statistical significance of relationships, or the Durbin-Watson statistic. As noted by Cox and Loomis (2001), authors of forecasting textbooks recommend the use of such criteria.

Statistical criteria are not appropriate for making comparisons among substantially different methods. They would be of little use to someone trying to choose between judgmental and quantitative methods, or among role playing, expert forecasts, and conjoint analysis. Statistical criteria are useful for selection only after the decision has been made about the general type of forecasting method, and even then their use has been confined primarily to quantitative methods.

Using statistical criteria for selection has other limitations. First, the criteria are usually absolute. Thus, the search for methods that are statistically significant can lead analysts to overlook other criteria and to ignore domain knowledge. Slovic and McPhillamy (1974) showed that when subjects were asked to choose between two alternatives, they often depended on a cue that was common to both alternatives and that was precisely measured, even when they recognized that this cue was irrelevant. Second, the rules are arbitrary in that they have no obvious relationship to decision making. They concern statistical significance, not practical significance.

Despite these problems, forecasters often use statistical criteria to select methods. This approach seems to be useful in some situations. For example, in extrapolation, statistical tests have helped forecasters to determine whether they should use seasonal factors and whether they should use a method that dampens trends. Judging from the M3-competition, statistical selection rules have been successfully employed for extrapolation. They can also help to select from among a set of econometric models (Allen and Fildes 2001).

RELATIVE TRACK RECORDS

The relative track record is the comparative performance of various methods as assessed by procedures that are systematic, unbiased, and reliable. It does *not* have to do with forecasting methods being used for a long time and people's satisfaction with them.

↳ **Compare the track records of various forecasting methods.**

Informal impressions often lead to different conclusions than those based on formal assessments. For example, most people believe that experts can predict changes in the stock market. Cowles (1933) examined 225 editorials by Hamilton, an editor for the *Wall Street Journal* who had gained a reputation as a successful forecaster. From 1902 to 1929, Hamilton forecasted 90 changes in the stock market; he was correct half the time and wrong the other half. Similar studies have followed in the stock market and related areas. Sherden's (1998, Chapter 4) analysis shows that the poor forecasting record of financial experts continues.

Assessing the track record is an appealing way to select methods because it eliminates the need to generalize from other research. The primary difficulty is that organizations seldom use good procedures for evaluating methods (Armstrong 2001a discusses these procedures). As a result, people have trouble distinguishing between a good track record and a good story.

Even if well designed, assessments of track records are based on the assumption that historical results can be generalized to the future. This can be risky, especially if the historical period has been stable, and the future situation is expected to be turbulent. To reduce risk, the analyst should assess the track record over a long time period. A longer history will provide more reliable estimates.

Few studies have been done on the value of using track records for selecting forecasting methods. The two studies that I found indicate that such assessments are useful.

Makridakis (1990) used the 111 series from the M-competition; these included annual, quarterly, and monthly data. He compared four methods: exponential smoothing with no trend, Holt's exponential smoothing with trend, damped trend exponential smoothing, and a long-memory autoregressive model. He deseasonalized the data when necessary. He compared the ex ante forecast errors on a holdout sample by using successive updating. For each series, he then used the model with the lowest MAPE for a given forecast horizon to forecast for a subsequent holdout sample. When methods were similar in forecast accuracy, Makridakis combined their forecasts. The accuracy of this procedure of selecting models based on horizon length accuracy was better than that achieved by the typical method (i.e., selecting a single model for all horizons). On average, it was a bit more accurate than equal-weights combining of forecasts.

Is it better to find the most accurate model for all series in a type (an aggregate selection strategy), or should one examine the track record for each series (individual selection)? To address this, Fildes (1989) examined data from a single organization. The series represented the number of telephone lines in use in each of 263 localities. He used two forecasting methods: Holt's exponential smoothing with an adjustment for large shifts, and a robust trend estimate. In making a robust trend estimate, one takes the median of the first differences (in this study an adjustment was also made for outliers). Fildes calibrated models on periods 1 to 24. He then used successive updating to make ex ante forecasts over periods 25 to 48. He used the error measures for this period as the basis for selection. He conducted a validation for periods 49 through 70. The strategies had similar accuracy for short-range forecasts (from one to six periods into the future). For longer-range forecasts (12- months-ahead), the error from aggregate selection was about six percent higher than that for individual selection. However, individual selection did no better than a combined forecast.

In the above comparisons, Makridakis and Fildes focused on accuracy. It would be useful to assess other criteria, such as the understandability and acceptability of each method. Another limitation is that these studies concern only extrapolation methods. I would expect track records to be especially useful when selecting from among substantially different methods.

PRINCIPLES FROM PUBLISHED RESEARCH

Assume that you needed to forecast personal computer sales in China over the next ten years. To determine which forecasting methods to use, you might use methods that have worked well in similar situations in the past. Having decided on this approach, you must consider: (1) How similar were the previous situations to the current one? (You would be unlikely to find comparative studies of forecasts of computer sales, much less computer sales in China), (2) Were the leading methods compared in earlier studies? (3) Were the evaluations unbiased? (4) Were the findings reliable? (5) Did these researchers examine the types of situations that might be encountered in the future? and (6) Did they compare enough forecasts?

Georgoff and Murdick (1986) made an early attempt to develop research-based guidelines for selection. They used 16 criteria to rate 20 methods. However, they cited only ten empirical studies. Because they were offering advice for a matrix with 320 cells, they depended primarily upon their opinions.

An extensive body of research is available for developing principles for selecting forecasting methods. The principles are relevant to the extent that the current situation is similar to those examined in the published research. Use of this approach is fairly simple and inexpensive.

General Principles

I examine some general principles from published research prior to discussing principles for various methods. The general principles are to use methods that are (1) structured, (2) quantitative, (3) causal, (4) and simple. I then examine how to match the forecasting methods to the situation.

☛ Use structured rather than unstructured forecasting methods

You cannot avoid judgment. However, when judgment is needed, you should use it in a structured way. For example, to forecast sales for a completely new product, you might use Delphi or intentions studies. Structured forecasting methods tend to be more accurate than unstructured ones. They are also easier to communicate and to replicate, and they aid learning, so the method can be improved over time.

☛ Use quantitative methods rather than judgmental methods, if enough data exist.

If no data exist, use judgmental methods. But when enough data exist, quantitative methods are expected to be more accurate. It is not always clear how much data is enough. This depends on the source, amount, relevance, variability, reliability, and validity of the data. The research to date offers little guidance. Studies such as the following would be useful; In a laboratory study on the time that groups took to assemble an Erector Set, Bailey and Gupta (1999) compared predictions made by 77 subjects against those from two quantitative learning-curve models. Bailey and Gupta provided data on the first two, four, six, or eight trials, and made predictions for the next three. Judgmental predictions were more accurate than quantitative methods given two or four observations. There was little difference given six observations. The quantitative approaches were more accurate than judgment when eight observations were available.

When sufficient data exist on the dependent variable and on explanatory variables, quantitative methods can be expected to be more accurate than judgmental methods. At worst, they seem to be as accurate. Few people believe this finding. There are some limiting conditions, but they are not that serious: First, the forecaster must be reasonably competent in using quantitative methods. Second, the methods should be fairly simple.

How can I make such a claim? The story goes back at least to Freyd (1925), who made a theoretical case that statistical procedures should be more accurate than judgmental procedures. Sarbin (1943) tested this in a study predicting the academic success of 162 college freshmen and found quantitative methods to be more accurate than judgmental forecasts. He thought that he had made a convincing case, and wrote:

“Any jury sitting in judgment on the case of clinical [judgmental] versus actuarial [statistical] methods must, on the basis of efficiency and economy, declare overwhelmingly in favor of the statistical method for predicting academic achievement.”

That was not the end of the story. Researchers questioned whether the findings would generalize to other situations. Paul Meehl published a series of influential studies on quantitative versus judgmental forecasting (e.g., Meehl 1954) and these extended Sarbin’s conclusion to other situations. In a more recent review, Grove and Meehl (1996) said it was difficult to find studies that showed judges to be more accurate than quantitative models. The results are consistent with those from judgmental bootstrapping (Armstrong 2001b) and expert systems (Collopy, Adya and Armstrong 2001). Despite much research evidence, practitioners still ignore the findings. As shown by Ahlburg (1991) and Dakin and Armstrong (1989), they even continue to rely on judgment for personnel predictions, the subject of much of this research.

The above studies concern cross-sectional predictions. Evidence from time series identifies some conditions under which judgmental methods are more accurate than quantitative methods. As with cross-sectional data, quantitative methods are likely to show the greatest accuracy when large changes are involved or many data are

available, but this is not so with few data. I summarized 27 empirical studies where few data were available and the expected changes were small (Armstrong 1985, pp. 393-400): Judgment was more accurate than quantitative methods in 17 studies, equally accurate in three studies, and less accurate in seven studies.

When you have enough data, then, use a quantitative method. This does not mean that one must avoid judgment. Indeed, you often need judgment as part of the process, for example, providing inputs or deciding which quantitative procedures to use.

☛ **Use causal rather than naive methods, especially if changes are expected to be large.**

Naive methods often give adequate results, and they are typically inexpensive. Thus, extrapolation methods may be appropriate for short-term inventory-control forecasts of products with long histories of stable demand. They are less effective in situations where there are substantial changes.

Causal methods, if well structured and simple, can be expected to be at least as accurate as naive methods. A summary of the evidence (Armstrong 1985, Exhibit 15-6), showed that causal methods were more accurate than naive methods in situations involving small changes in nine comparative studies, the same in six, and less accurate in six. But, for long-term forecasts (large changes), all seven studies showed that causal methods were more accurate. Allen and Fildes (2001) extended the analysis and found that causal methods were more accurate than extrapolations for short-term forecasts for 34 studies and less accurate for 21 (using the *ex ante* forecast error for “short” and “short/medium” from their Table A4 found at the forecasting principles website, hops.wharton.upenn.edu/forecast). For their “medium” and “medium-long” forecasts, causal methods were more accurate for 58 studies and less accurate for 20.

Does this principle hold up in practice? Bretschneider et al. (1989) obtained information on 106 sales tax forecasts and 74 total revenue forecasts from state governments in the U.S. These were one-year ahead annual forecasts for 1975-85 from 28 states that responded to a survey. States using quantitative methods had smaller errors than states using judgmental methods.

☛ **Use simple methods unless substantial evidence exists that complexity helps.**

Use simple methods unless a strong case can be made for complexity. One of the most enduring and useful conclusions from research on forecasting is that simple methods are generally as accurate as complex methods. Evidence relevant to the issue of simplicity comes from studies of judgment (Armstrong 1985, pp. 96-105), extrapolation (Armstrong 1984, Makridakis et al. 1982, and Schnaars 1984), and econometric methods (Allen and Fildes 2001). Simplicity also aids decision makers’ understanding and implementation, reduces the likelihood of mistakes, and is less expensive.

Simplicity in an econometric model would mean a small number of causal variables and a functional form that is linear in its parameters (e.g., an additive model or a log-log model). For extrapolation, it might mean nothing more complex than exponential smoothing with seasonally adjusted data. For role-playing, it would mean brief sessions based on short role descriptions. An operational definition of *simple* is that the approach can be explained to a manager so clearly that he could then explain it to others.

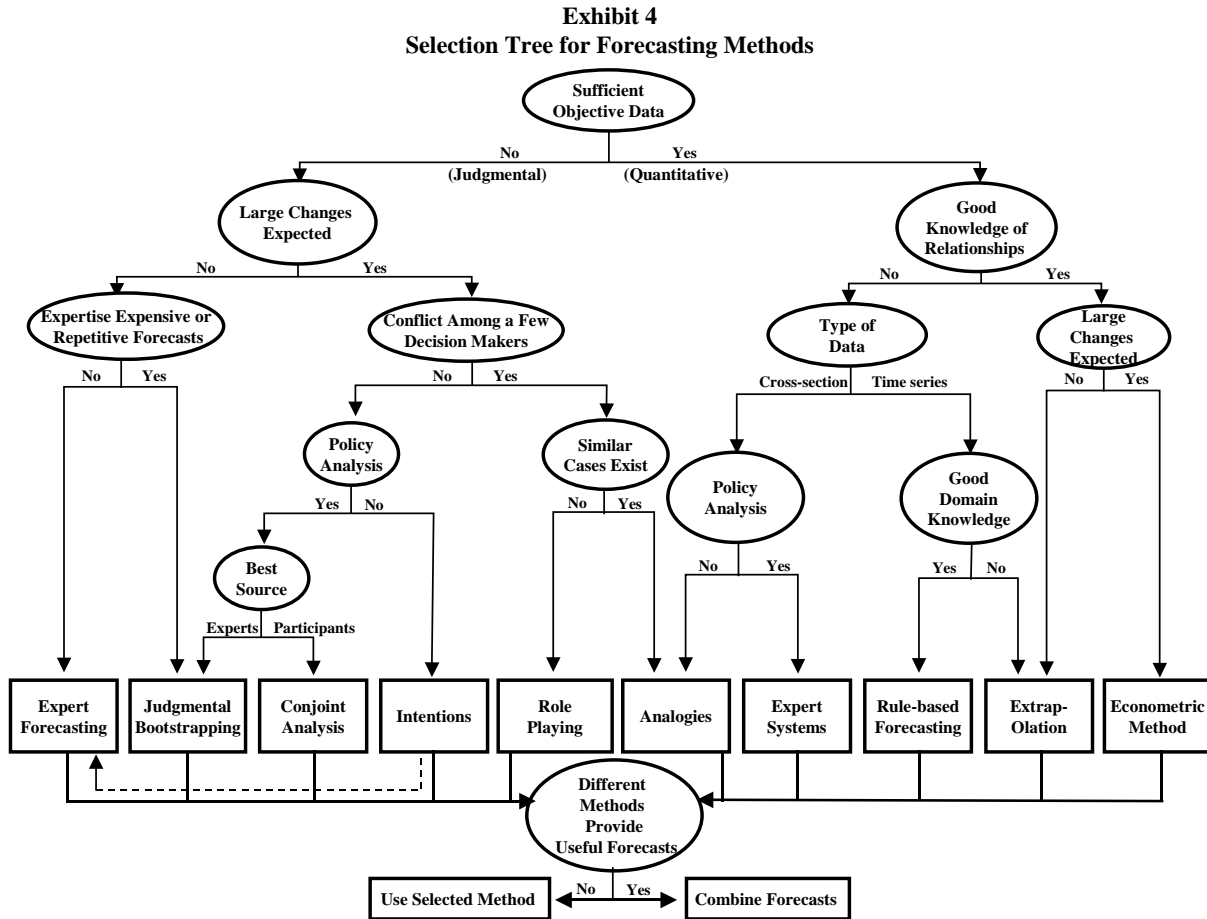
In his review of population forecasting, Smith (1997) concluded that simple methods are as accurate as complex methods. In their review of research on tourism forecasting, Witt and Witt (1995) concluded that the naive (no-change) model is typically more accurate than other procedures, such as commercially produced econometric models. The value of simplicity shows up in practice; in a survey on the accuracy of U.S. government revenue forecasts, states that used simple econometric methods reported substantially lower MAPEs than those that used complex econometric methods (Bretschneider et al. 1989).

Nevertheless, some complexity may help when the forecaster has good knowledge of the situation. Simple econometric methods are often more accurate than extrapolations (Allen and Fildes 2001, Tables A5, A6 and A7 on the forecasting principles website). Decomposed judgments are often more accurate than global judgments (MacGregor 2001). Exponential smoothing of trends is often more accurate than naive forecasts. In fact, many forecasting principles call for added complexity. That said, forecasters often use overly complex methods.

Complexity improves the ability to fit historical data (and it probably helps to get papers published), but it often harms forecast accuracy.

☛ **Match the forecasting methods to the situation.**

The above general principles were used, along with prior research, to develop more specific guidelines for selecting methods based on the situation. They are described here, along with evidence, following the flow chart in Exhibit 4.



Judgmental Methods

Starting with the judgmental side of the selection tree, the discussion proceeds downward, and then from left to right.

The selection of judgmental procedures depends on whether substantial deviations from a simple historical projection are expected over the forecast horizon. When predicting for cross-sectional data, the selection of a method depends on whether large differences are expected among the elements to be forecast; for example, the performances of players selected by professional sports teams will vary enormously.

Small Changes: If the expected changes are not large, methods are likely to differ little in accuracy. Also for infrequent forecasts, *expert forecasts*, which can be tailored to the situation and prepared quickly, may be sufficient.

If many forecasts are needed, expert forecasts are likely to be too expensive. For example, the demand for items in a sales catalogue or the success of job candidates require many forecasts. *Judgmental bootstrapping* is appropriate in such cases. It can provide forecasts that are less expensive than those based on judgment because it

applies the experts' procedures in a mechanical way. In addition, bootstrapping will provide improvements in accuracy.

Large Changes with No Conflicts: In some situations, you may expect large changes (moving to the right in Exhibit 4). If decision makers in the situation are not in conflict, you can obtain forecasts from experts or participants.

When decision makers need forecasts to examine different policies, they can obtain them from experts and participants. Judgmental bootstrapping and conjoint analysis are well suited for this.

A company planning to sell computers in China might need forecasts to make decisions on pricing, advertising, and design for a new product. *Judgmental bootstrapping* provides a low-cost way to examine a wide range of policies. For example, experts could make forecasts for about 20 different marketing plans constructed according to an experimental design. The model could then be used to predict responses to still other plans. The key assumption behind judgmental bootstrapping is that the experts who provide inputs to the model understand the situation. Judgmental bootstrapping is superior to expert forecasts in terms of accuracy (Armstrong 2001b). It also provides consistent forecasts, which helps in comparing alternative policy options. In addition, judgmental bootstrapping offers an opportunity to evaluate some policy variables that cannot be examined by conjoint analysis. For example, what would happen to sales if a firm increased advertising for a particular product? An expert could assess this, but not a prospective customer.

When experts lack experience to judge how customers will respond, it may help to seek information from potential customers. *Conjoint analysis* can be used to develop a forecasting model based on how consumers respond to alternative offerings. If the proposed product or service is new, they might not know how they would respond. But if the alternatives are described realistically, they can probably predict their actions better than anyone else can. Realistic descriptions can be done at a low cost (Armstrong and Overton 1971). As Wittink and Bergestuen (2001) discuss, conjoint analysis offers a consistent way to evaluate alternatives and it improves forecast accuracy. However, given the need for large samples, this can be expensive.

For important forecasts, you can use both judgmental bootstrapping and conjoint analysis. Their forecasts for policy options might differ, in which case you gain information about forecast uncertainty. A combination of forecasts from the two methods would likely improve accuracy and reduce the risk of large errors.

If there is no need to forecast for alternative policies, I recommend *intentions* studies. Present the issue and ask people how they would respond. For example, this approach could be used to predict the vote for a referendum to reduce taxes. Or, as Fullerton and Kinnaman (1996) did, it could be used to predict how people would respond to a plan to charge residents for each bag of garbage.

Expert forecasts can also be used to assess a proposed policy change. For example, in one project we asked a sample of potential customers about their intentions to subscribe to a proposed urban mass transit system known as the Minicar Mass Transit System (Armstrong and Overton 1971). As an alternative approach, we could have described the system's design and marketing plan to a group of, say, six mass-transportation experts, and asked them to predict the percentage of the target market that would subscribe to the service over the next year. Such a survey of experts would have been faster and cheaper than the intentions study.

Lemert (1986) asked 58 political experts to predict the outcomes of two referendums on the 1982 Oregon ballot. One dealt with land-use planning and the other with property-tax limitations. Although the vote was nearly tied in each case, the experts were usually correct (73.2% were correct on the first and 89.5% on the second). Moreover, the mean prediction of the experts was close to the actual vote (off by 1.4% on the first issue and by 0.3% on the second). But when Lemert obtained predictions from 283 *voters*, fewer were correct (63.3% and 61.4% respectively). Of those who voted "yes" (averaging across the two issues), 70% expected the referendum would be passed. Of those who voted "no," 25% thought it would be passed. This study demonstrates that unbiased experts are more accurate than participants in predicting the behavior of other people.

Some confusion exists about the use of intentions and expert opinions. Because experts forecast the behavior of many people, you need only a few experts. Lewis-Beck and Tien (1999) exhibited this confusion. They

compared the results of their surveys of voters' intentions against another survey that asked voters to predict who would win. In the latter case, the researchers used voters as experts. This was a poor strategy because most voters lack adequate knowledge about others, and they are biased in favor of their candidates. Lewis-Beck and Tien compensated for such bias by selecting representative probability samples. This required samples of from 1,000 to 2,000 each year, whereas a sample of ten unbiased political experts probably would have been adequate.

Large Changes with Conflicts: When considering situations with large changes, it is difficult to find relevant analogies. For example, when Fred Smith started FedEx in the mid-1970s, the U.S. Post Office charged \$1.50 for mailing a special delivery letter. FedEx planned to provide a faster and more reliable service for \$12.50. *Role-playing* would have been useful to forecast competitors' behavior. People could play roles as key executives at FedEx, the U.S. Post Office, and perhaps UPS. They would be asked to respond to various plans FedEx was considering. Role playing is more accurate than expert forecasts in situations in which two parties are in conflict with each other (Armstrong 2001b).

Analogies can be useful. For instance, in trying to predict how legalization of drugs would affect the number of users and crime rates, look to studies of the prohibition of alcohol in the U.S. and other countries. To predict the sales of brand-name drugs a year after the introduction of generic drugs, generalize from previous situations; according to the *Wall Street Journal* (Feb. 20, 1998, p. B5), brand-name drugs lose about 80% of their dollar sales.

It can help to merely *think* about analogies and to consider how the current situation relates to them. Cooper, Woo and Dunkelberg (1988) asked 2,994 new entrepreneurs to estimate their perceived chances of success. Eighty-one percent of them thought their odds were better than seven in ten. Interestingly, those who were poorly prepared to run a business were just as optimistic as those who were better prepared. But when they were asked "What are the odds of any business like yours succeeding," only 39% thought the odds were better than seven in ten. Based on studies reviewed by Cooper, Woo and Dunkelberg, even this estimate exceeds the historical success rate of entrepreneurs. Still, thinking about analogies could have led these entrepreneurs to more accurate forecasts.

Quantitative Methods

When you have enough objective data to use quantitative methods (the right-hand side of Exhibit 4), you may or may not have good prior knowledge about future relationships. When you do not, the selection of an approach depends on whether you have cross-sectional or time-series data.

Poor Knowledge of Relationships and Cross-sectional Data: If you lack knowledge of expected relationships, and have cross-sectional data, ask whether you need to compare alternative policies. If not, experts can use *analogies* as the basis for forecasts.

Use unbiased procedures to select a large sample of analogies. For example, in trying to predict whether a campaign to introduce water fluoridation in a particular community in New Zealand will succeed, one could analyze the many analogous cases in the U.S. This advice is often ignored. Consider the following case. Stewart and Leschine (1986) discussed the use of analogies for the decision to establish an oil refinery in Eastport, Maine. The Environmental Protection Agency had not used worldwide estimates of tanker spills, but instead relied on a single analogy (Milford Haven in the U.K.) believing that it was a comparable site. The use of a single site is unreliable and prone to bias. Analysts should have rated all ports for similarity without knowledge of their oil spill rates, selected some of the most similar, and then examined spill rates.

Information from analogies can reduce the effects of potential biases because analogies provide objective evidence. This was illustrated in Kahneman and Lovallo (1993). Kahneman had worked with a small team of academics to design a new judgmental decision making curriculum for Israeli high schools. He circulated slips of paper and asked each team member to predict the number of months it would take them to prepare a proposal for the Ministry of Education. The estimates ranged from 18 to 30 months. Kahneman then turned to a member of the team who had considerable experience developing new curricula. He asked him to think of analogous projects. The man recalled that about 40% of the teams eventually gave up. Of those that completed the task, he said, none did so in

less than seven years. Furthermore, he thought that the present team was probably below average in terms of resources and potential. As it turned out, it took them eight years to finish the project.

Experiments by Buehler, Griffin and Ross (1994) supported Kahneman and Lovallo's illustration. Their subjects made more accurate predictions of the time they would take to do a computer assignment when they described analogous tasks they had solved previously. Without the analogies, there were overly optimistic. The subjects were even more accurate when they described how the current task related to analogous cases they had experienced.

If no suitable analogies can be found, you might try to create them by conducting field or laboratory experiments. Field experiments are more realistic. As a result, they are generally thought to provide more valid forecasts. They are widely used in test marketing new products to predict future sales. On the negative side, field experiments are subject to many threats to validity. Competitors may respond in test markets so as to distort the forecasts and environmental changes may affect test results.

Laboratory experiments offer more control. Despite claims that they lack external validity and suffer from what reviewers delightfully refer to as "demand effects" (subjects just responding to the demand of the experiment), laboratory experiments are often useful for forecasting. More generally, Locke (1986), using a series of studies in organizational behavior, showed that findings from laboratory experiments were generally similar to those from field experiments.

The key is to design experiments, whether laboratory or field, so that they match the forecasting situation reasonably well. For example, in a lab experiment designed to estimate price elasticities, Wright and Gendall (1999) showed that it was important to at least provide a picture of the product and to consider only responses from potential purchasers. Previous studies, in which researchers had not done this often produced inaccurate estimates. Conjoint studies sometimes fail to provide adequate illustrations. The Internet provides a low-cost way to provide realistic descriptions. Dahan and Srinivasan (2000), in a study of 11 different bicycle pumps, found that a web-based descriptions were similar to physical prototypes in predicting market share. Web-based designs are much less expensive than physical prototypes.

When people need to compare alternative policies, *expert systems* should be considered. They are especially useful when the situation is complex and experts differ in their ability to forecast. An expert system should be based on the processes used by those thought to be the best experts.

Judgmental bootstrapping is also relevant for comparing policies. You can infer rules by regressing the experts' predictions against actual data. Alternatively, you can infer the experts' rules by asking them to make predictions for fictitious (but realistic) cases. This latter approach is appropriate when historical values do not have large variations and when the historical variations are not independent of one another.

The choice between expert systems and judgmental bootstrapping is likely to be based on costs and complexity. Judgmental bootstrapping is less expensive, but it requires a great deal of simplification. If complexity is needed and you have excellent domain knowledge, expert systems might enable the description of a well-structured set of conditions that can improve forecast accuracy (Collopy, Adya and Armstrong 2001).

Poor Knowledge of Relationships and Time Series Data: Although you may lack good prior knowledge of relationships, you may be able to obtain specific knowledge about a situation. For example, a manager may know a lot about a product and this might help in preparing a sales forecast.

If good domain knowledge is available, *rule-based forecasting* is appropriate. Although it is more costly than extrapolation, RBF tends to improve accuracy (versus pure extrapolation) because it uses domain knowledge and because the rules tailor the extrapolation method to the situation (Armstrong, Adya and Collopy 2001).

Rule-based forecasting might also be appropriate if domain knowledge is not available. This is because it applies guidelines from prior research. However, little research has been done to test this proposition. Still, its accuracy was competitive with the best of well-established software programs when used for annual data in the M3-competition (Adya, Armstrong, Collopy and Kennedy 2000).

Extrapolation of time series is a sensible option if domain knowledge is lacking, the series is stable and many forecasts are needed. These conditions often apply to forecasting for inventory control. I suspect, however, that people have useful domain knowledge in most situations.

Good Knowledge of Relationships and Small Changes (far right of Exhibit 4): Knowledge of relationships might be based on the judgment of experts who have received good feedback in previous comparable situations or on the results of empirical studies. For example, in trying to predict the effects of alternative marketing plans for a product, one might rely on the many studies of price and advertising elasticities, such as those summarized by Tellis (1988), Assmus, Farley and Lehmann (1984), and Sethuraman and Tellis (1991).

When small changes are expected, knowledge about relationships is not of much value. Difficulties in measurement and in forecasting changes in the causal variables are likely to negate the value of the additional information. Thus, studies involving short-term forecasting show that *extrapolation* methods (which ignore causal information) are often as accurate as econometric methods (Allen and Fildes 2001).

Expert forecasts can be expected to do well in these situations. In line with this expectation, Braun and Yaniv (1992) found that economists were more accurate than quantitative models in estimating the level at time t_0 (when changes are small), as accurate in forecasting one-quarter-ahead forecasts, but less accurate for four-quarters-ahead.

Good Knowledge of Relationships and Large Changes: Use *econometric* methods when large changes are expected. The evidence summarized by Allen and Fildes (2001) supports this advice.

In my study of the photographic market (Armstrong 1985, p. 411), where there was good knowledge of relationships, I made six-year backcasts of camera sales, using the data from 1965 through 1960 to backcast for 1954. The data were put into three groups: six countries with moderate changes in sales, five with large changes, and six with very large changes. For the six countries with moderate changes, the errors from an econometric model averaged 81% of errors from a combined forecast based on no trend, the trend for that country, and the trend for all 17 countries. For five countries with large changes, the errors averaged 73% of the combined extrapolations, and for five countries with very large changes, they were 32%. As hypothesized, econometric methods were relatively more accurate than trend extrapolation when change was largest. This study was limited because it used actual changes (not expected changes) in the dependent variable.

A study of the air travel market (Armstrong and Grohman 1972) also showed the value of econometric methods to be greater when large changes were expected. This was an ideal situation for econometric models because there was good prior knowledge and ample data. Analysts at the Federal Aviation Agency (FAA) had published judgmental forecasts for the U.S. market. They had access to all of the knowledge and used quantitative methods as inputs to their judgmental forecasts. Armstrong and Grohman (1972) examined forecasts for 1963 to 1968 using successive updating. In this case, the expected change was based simply on the length of the forecast horizon; more change being expected in the long-run. As shown in Exhibit 5, the econometric forecasts were more accurate than the FAA's judgmental forecasts and this gain became greater as the horizon increased.

Exhibit 5
Accuracy of Econometric Models in an Ideal Situation: U.S. Air Travel

Forecast Horizon (years)	Number of Forecasts	Mean Absolute Percentage Errors		
		Judgment by FAA	Econometric Model	Error Reduction
1	6	6.8	4.2	2.6
2	5	15.6	6.8	8.8
3	4	25.1	7.3	17.8
4	3	34.1	9.8	24.3
5	2	42.1	6.2	35.9
6	1	45.0	0.7	44.3

Besides improving accuracy, econometric methods allow you to compare alternative policies. Furthermore, they can be improved as you gain knowledge about the situation.

IMPLICATIONS FOR PRACTITIONERS

First consider what not to do. Do not select methods based on convenience, except for stable situations with little change and where accuracy is not critical.

The popularity of a method does not indicate its effectiveness. It provides little information about the performance of the methods and about the situations in which they are used. Furthermore, forecasters may overlook relevant methods.

Structured judgment is valuable, especially if ratings by forecasting experts are used. First develop criteria, and then ask experts for formal (written) ratings of how various methods meet those criteria.

Statistical criteria are important and should become more useful as researchers examine how they relate to accuracy. Still, some statistical criteria are irrelevant or misleading. Furthermore, statistical criteria may lead analysts to overlook relevant criteria.

When large changes are expected and errors have serious consequences, you can assess the track record of leading forecasting methods in the given situation. While useful and convincing, comparing the accuracy of various methods is expensive and time-consuming.

Drawing upon extensive research, we developed guidelines to help practitioners decide which methods are appropriate for their situations. Through these guidelines, one can select methods rapidly and inexpensively. If a number of methods are promising, use them and combine their forecasts (Armstrong 2001c).

IMPLICATIONS FOR RESEARCHERS

To assess market popularity, you would need to learn about a method's performance relative to other methods. The type of research by Bretschneider et al. (1989) is promising. They used survey data from state government agencies. Respondents described their revenue forecasting methods and reported actual values. Bretschneider et al. correlated the forecasting methods they used with their forecast errors. Studies of market popularity should also identify the conditions (e.g., were large changes expected? Was there high uncertainty?). The survey should go beyond "use" to consider "satisfaction" and "performance."

Do structured procedures help analysts to select good forecasting methods? The evidence I have cited did not come from studies on forecasting, so it would be worthwhile to directly examine the value of structured

procedures for selecting forecasting methods. For example, you could use situations for which researchers have identified the best methods, but you would not reveal this to the forecasters. The question is whether, given a description of the current situation, forecasters who follow a structured approach would make a better selection of forecasting methods than forecasters with similar experience who do not use a structured approach.

Statistical criteria have been assumed to be useful. However, little research has been done to examine the effectiveness of statistical criteria. Comparative studies are needed. The M-competitions do not meet the need because the various methods differ in many ways. Thus one cannot determine which aspects of the methods are effective under various conditions.

I was able to find only two studies that assessed the use of relative track records for selection. This should be a fertile area for further research.

Research that contributes to the development and refinement of guidelines for selection is always useful. Such findings can be easily applied to the selection of forecasting methods if the conditions are well defined.

CONCLUSIONS

I described six approaches for selecting forecasting methods. Convenience and market popularity, while often used, are not recommended. Structured judgment, statistical criteria, and track records can all help in selecting and can be used in conjunction with one another. Guidelines from prior research are particularly useful for selecting forecasting methods. They offer a low-cost way to benefit from findings based on expert judgments and on over half a century of research on forecasting.

REFERENCES

- Adya, M., J. S. Armstrong, F. Collopy & M. Kennedy (2000), "An application of rule-based forecasting to a situation lacking domain knowledge," *International Journal of Forecasting* (forthcoming).
- Ahlburg, D. (1991), "Predicting the job performance of managers: What do the experts know?" *International Journal of Forecasting*, 7, 467-472.
- Allen, G. & R. Fildes (2001), "Econometric forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S. (1984), "Forecasting by extrapolation: Conclusions from twenty-five years of research," (with commentary), *Interfaces*, 14 (Nov.-Dec.), 52-66. Full text at hops.wharton.upenn.edu/forecast.
- Armstrong, J. S. (1985), *Long-Range Forecasting: From Crystal Ball to Computer*. New York: John Wiley. Full text at hops.wharton.upenn.edu/forecast.
- Armstrong, J. S. (2001a), "Evaluating forecasting methods," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S. (2001b), "Judgmental bootstrapping: Inferring experts' rules for forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S. (2001c), "Combining forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S. (2001d), "Standards and practices for forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.

- Armstrong, J. S., M. Adya & Collopy, F. (2001), "Rule-based forecasting: Using judgment in time-series extrapolation," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA. Kluwer Academic Publishers.
- Armstrong, J. S., R. Brodie & S. McIntyre (1987), "Forecasting methods for marketing," *International Journal of Forecasting*, 3, 355-376. Full text at hops.wharton.upenn.edu/forecast.
- Armstrong, J. S. & M. Grohman (1972), "A comparative study of methods for long-range market forecasting," *Management Science*, 19, 211-221. Full text at hops.wharton.upenn.edu/forecast.
- Armstrong, J. S. & T. Overton (1971), "Brief vs. comprehensive descriptions in measuring intentions to purchase," *Journal of Marketing Research*, 8, 114-117. Full text at hops.wharton.upenn.edu/forecast.
- Assmus, G., J. U. Farley & D. R. Lehmann (1984), "How advertising affects sales: A meta-analysis of econometric results," *Journal of Marketing Research*, 21, 65-74.
- Bailey, C. D. & S. Gupta (1999), "Judgment in learning-curve forecasting: A laboratory study," *Journal of Forecasting*, 18, 39-57.
- Braun, P. A. & I. Yaniv (1992), "A case study of expert judgment: Economists' probabilities versus base-rate model forecasts," *Journal of Behavioral Decision Making*, 5, 217-231.
- Bretschneider, S. I., W. L. Gorr, G. Grizzle & E. Klay (1989), "Political and organizational influences on the accuracy of forecasting state government revenues," *International Journal of Forecasting*, 5, 307-319.
- Buehler, R., D. Griffin & M. Ross (1994), "Exploring the 'planning fallacy': Why people underestimate their task completion times," *Journal of Personality and Social Psychology*, 67, 366-381.
- Campion, M. A., D. K. Palmer & J. E. Campion (1997), "A review of structure in the selection interview," *Personnel Psychology*, 50, 655-701.
- Carbone, R. & J. S. Armstrong (1982), "Evaluation of extrapolation forecasting methods: Results of a survey of academicians and practitioners," *Journal of Forecasting*, 1, 215-217. Full text at hops.wharton.upenn.edu/forecast.
- Chambers, J. C., S. Mullick & D. D. Smith (1971), "How to choose the right forecasting technique," *Harvard Business Review*, 49, 45-71.
- Chambers, J. C., S. Mullick & D. D. Smith (1974), *An Executive's Guide to Forecasting*. New York: John Wiley.
- Collopy, F., M. Adya & S. Armstrong (2001), "Expert systems for forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA. Kluwer Academic Publishers.
- Cooper, A., C. Woo & W. Dunkelberg (1988), "Entrepreneurs' perceived chances for success," *Journal of Business Venturing*, 3, 97-108.
- Cowles, A. (1933), "Can stock market forecasters forecast?" *Econometrica*, 1, 309-324.
- Cox, J. E., Jr. & D. G. Loomis (2001), "Diffusion of forecasting principles: An assessment of books relevant to forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Dahan, E. & V. Srinivasen (2000), "The predictive power of internet-based product concept testing using visual depiction and animation," *Journal of Product Innovation Management*, 17, 99-109.
- Dakin, S. & J. S. Armstrong (1989), "Predicting job performance: A comparison of expert opinion and research findings," *International Journal of Forecasting*, 5, 187-194. Full text at hops.wharton.upenn.edu/forecast.

- Darlymple, D. J. (1987), "Sales forecasting practices: Results from a United States survey," *International Journal of Forecasting*, 3, 379-391.
- Fildes, R. (1989), "Evaluation of aggregate and individual forecast method selection rules," *Management Science*, 35, 1056-1065.
- Frank, H. A. & J. McCollough (1992) "Municipal forecasting practice: 'Demand' and 'supply' side perspectives," *International Journal of Public Administration*, 15, 1669-1696.
- Freyd, M. (1925), "The statistical viewpoint in vocational selection," *Journal of Applied Psychology*, 9, 349-356.
- Fullerton, D. & T. C. Kinnaman (1996), "Household responses to pricing garbage by the bag," *American Economic Review*, 86, 971-984.
- Georgoff, D. M. & R. G. Murdick (1986), "Manager's guide to forecasting," *Harvard Business Review*, (January-February), 110-120.
- Grove, W. M. & P. E. Meehl (1996), "Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy," *Psychology, Public Policy and Law*, 2, 293-323.
- Kahneman, D. & D. Lovallo (1993), "Timid choices and bold forecasts: A cognitive perspective on risk taking," *Management Science*, 39, 17-31.
- Lemert, J. B. (1986), "Picking the winners: Politician vs. voter predictions of two controversial ballot measures," *Public Opinion Quarterly*, 50, 208-221.
- Lewis-Beck, M. S. & C. Tien (1999), "Voters as forecasters: A micromodel of election prediction," *International Journal of Forecasting*, 15, 175-184.
- Locke, E. A. (1986), *Generalizing from Laboratory to Field Settings*. Lexington, MA: Lexington Books.
- MacGregor, D. G. (2001), "Decomposition for judgmental forecasting an estimation," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Mahmoud, E., G. Rice & N. Malhotra (1986), "Emerging issues in sales forecasting on decision support systems," *Journal of the Academy of Marketing Science*, 16, 47-61.
- Makridakis, S. (1990), "Sliding simulation: A new approach to time series forecasting," *Management Science*, 36, 505-512.
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.
- Meehl, P. E. (1954), *Clinical vs. Statistical Prediction*. Minneapolis: University of Minnesota Press.
- Mentzer, J. T. & J. E. Cox, Jr. (1984), "Familiarity, application, and performance of sales forecasting techniques," *Journal of Forecasting*, 3, 27-36
- Mentzer, J. T. & K. B. Kahn (1995), "Forecasting technique familiarity, satisfaction, usage, and application," *Journal of Forecasting*, 14, 465-476.
- Rhyne, D. M. (1989), "Forecasting systems in managing hospital services demand: A review of utility," *Socio-economic Planning Sciences*, 23, 115-123.

- Rowe, G. & G. Wright (2001), "Expert opinions in forecasting: Role of the Delphi technique," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Sanders, N. R. & K. B. Manrodt (1994), "Forecasting practices in U. S. corporations: Survey results," *Interfaces*, 24 (2), 92-100.
- Sarbin, T. R. (1943), "A contribution to the study of actuarial and individual methods of prediction," *American Journal of Sociology*, 48, 593-602.
- Schnaars, S. P. (1984), "Situational factors affecting forecast accuracy," *Journal of Marketing Research*, 21, 290-297.
- Sethuraman, R. & G. J. Tellis (1991), "An analysis of the tradeoff between advertising and price discounting," *Journal of marketing Research*, 28, 160-174.
- Sherden, W. A. (1998), *The Fortune Sellers*. New York: John Wiley.
- Slovic, P. & D. J. McPhillamy (1974), "Dimensional commensurability and cue utilization in comparative judgment," *Organizational Behavior and Human Performance*, 11, 172-194.
- Smith, S. K. (1997), "Further thoughts on simplicity and complexity in population projection models," *International Journal of Forecasting*, 13, 557-565.
- Stewart, T. (2001), "Improving reliability of judgmental forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Stewart, T. R. & T. M. Leschine (1986), "Judgment and analysis in oil spill risk assessment," *Risk Analysis* 6 (3), 305-315.
- Tellis, G. J. (1988), "The price elasticity of selective demand: A meta-analysis of econometric models of sales," *Journal of Marketing Research*, 25, 331-341.
- Witt, S. F. & C. A. Witt (1995), "Forecasting tourism demand: A review of empirical research," *International Journal of Forecasting*, 11, 447-475.
- Wittink, D. R. & T. Bergestuen (2001), "Forecasting with conjoint analysis," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- Wright, M. & P. Gendall (1999), "Making survey-based price experiments more accurate," *Journal of the Market Research Society*, 41, (2) 245-249.
- Yokum, T. & J. S. Armstrong (1995), "Beyond accuracy: Comparison of criteria used to select forecasting methods," *International Journal of Forecasting*, 11, 591-597. Full text at hops.wharton.upenn.edu/forecast.

Acknowledgments: P. Geoffrey Allen, William Ascher, Lawrence D. Brown, Derek Bunn, Fred Collopy, Stephen A. DeLurgio and Robert Fildes provided useful comments on early drafts. Jennifer L. Armstrong, Raphael Austin, Ling Qiu and Mariam Rafi made editorial revisions.