DEMAND FORECASTING: EVIDENCE-BASED METHODS

Forthcoming in the *Oxford Handbook in Managerial Economics*
Christopher R. Thomas and William F. Shughart II (Eds.)
Subject to further revisions
File: Demandforecasting-17-August-2011-clean.docx

17 August 2011

J. Scott Armstrong
The Wharton School, University of Pennsylvania
747 Huntsman, Philadelphia, PA 19104, U.S.A.
T: +1 610 622 6480 F: +1 215 898 2534 armstrong@wharton.upenn.edu

Kesten C. Green
International Graduate School of Business, University of South Australia
City West Campus, North Terrace, Adelaide, SA 5000, Australia
T: +61 8 8302 9097 F: +61 8 8302 0709 kesten.green@unisa.edu.au

words in body 10,053 (requested range was 6,000 to 9,000)

ABSTRACT

We reviewed the evidence-based literature related to the relative accuracy of alternative methods for forecasting demand. The findings yield conclusions that differ substantially from current practice. For problems where there are insufficient data, where one must rely on judgment. The key with judgment is to impose structure with methods such as surveys of intentions or expectations, judgmental bootstrapping, structured analogies, and simulated interaction. Avoid methods that lack evidence on efficacy such as intuition, unstructured meetings, and focus groups. Given ample data, use quantitative methods including extrapolation, quantitative analogies, rule-based forecasting, and causal methods. Among causal methods, econometric methods are useful given good theory, and few key variables. Index models are useful for selection problems when there are many variables and much knowledge about the situation. Use structured procedures to incorporate managers' domain knowledge into forecasts from quantitative methods where the knowledge would otherwise be overlooked, but avoid unstructured revisions. Methods for combining forecasts, including prediction markets and Delphi, improve accuracy. Do not use complex methods; they do not improve accuracy and the added complexity can cause forecasters to overlook errors and to apply methods improperly. We do not recommend complex econometric methods. Avoid quantitative methods that have not been properly validated and those that do not use domain knowledge; among these we include neural nets, stepwise regression, and data mining. Given that many organizations use the methods we reject and few use the methods we recommend, there are many opportunities to improve forecasting and decision-making.

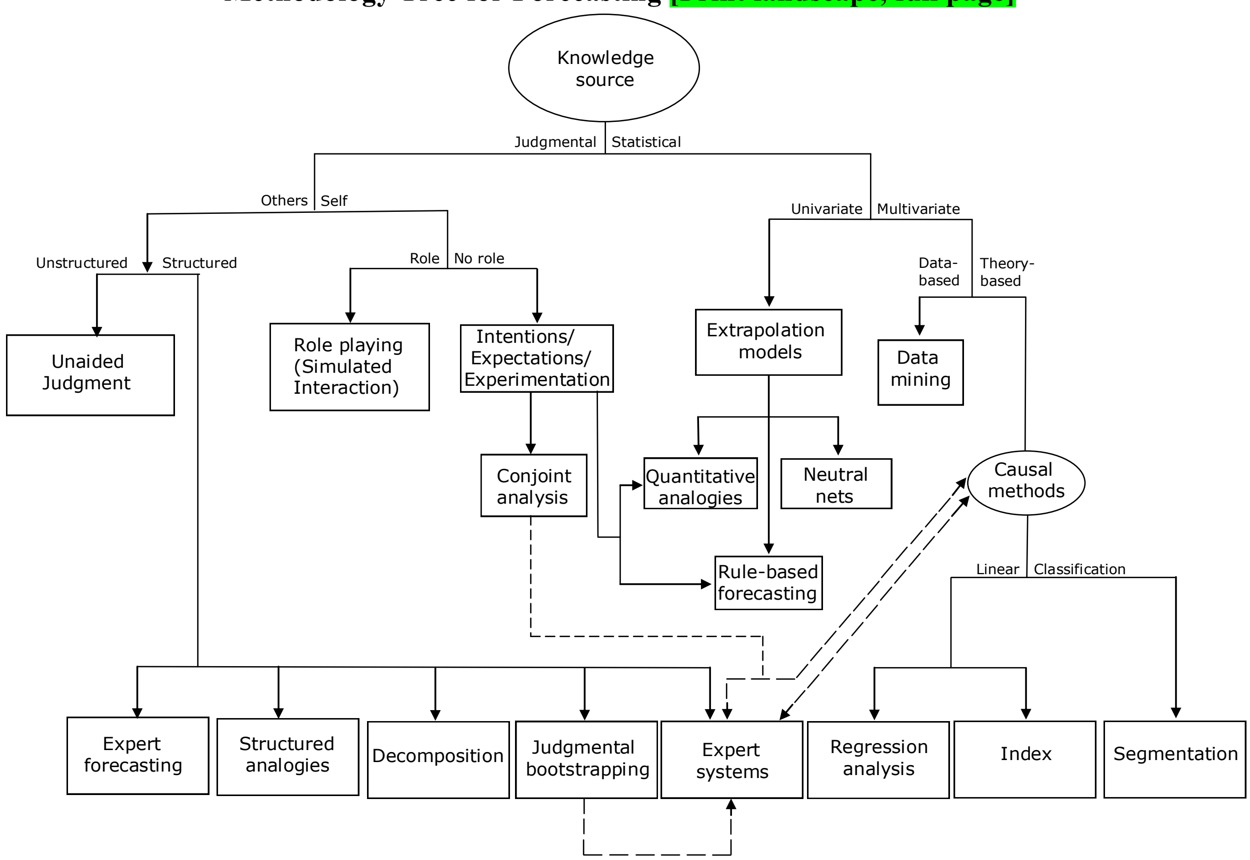
Keywords: competitor behaviour, data mining, Delphi, expectations, game theory, intentions, market share, index models, market size, forecast accuracy, forecasting methodology, prediction markets, sales forecasting, simulated interaction, stepwise regression, structured analogies.

Demand forecasting asks "how much can be sold given the situation and the marketing program?" The situation includes the broader economy, infrastructure, the social environment, the legal framework, the market, actions by the firm, actions by those offering competing and complementary products, and actions by others such as unions and lobby groups.

Overview of possible methods

In this section we provide brief descriptions of the types of forecasting methods that might be used to forecast demand and the evidence that is available on their use. The forecasting methods and the relationships among them are shown in Figure 1, the Methodology Tree for Forecasting. The primary distinction is between methods that rely on judgement and those that estimate relationships using quantitative data.

Figure 1
Methodology Tree for Forecasting [Print landscape, full page]



Methods Based on Judgment

Unaided judgment

Most important demand-related forecasts in organizations are made using unaided judgment. By "unaided" we mean judgment that does not use of evidence-based procedures. Such forecasts might include those for the sales of a new product, the effects of a change in design, pricing, or advertising, or how competitors would respond. Forecasts by experts using their unaided judgment are most likely to be accurate when the situation

- is similar to others that the experts have made forecasts about,
- involves relationships that are simple and well understood (e.g., demand goes up when prices go down),
- is unlikely to be effected by large changes, and
- does not involve conflict,

and when the experts

- are unbiased,
- possess information that others do not, and
- receive accurate, timely, and well-summarized feedback about their forecasts.

Regrettably, unaided judgement is often used when the above conditions do not hold. For example, these conditions do not apply to political and economic forecasting. Tetlock's (2005) study of more than 82,000 forecasts made over 20 years by 284 experts in politics and economics found that they were little more accurate than those made by non-experts and that they were less accurate than forecasts from simple models. His findings are consistent with those from research in other fields that involve uncertainty and complex problems—such as is the case for many problems related to demand forecasting.

Prediction markets

Prediction markets, which are also known as betting markets, information markets, and futures markets, have a long history. Between the end of the U.S. Civil War and World War II, well-organized markets for betting on U.S. presidential elections correctly picked the winner in every election but 1916; also, they were successful in identifying those elections that would be close (Rhode and Strumpf 2004). More recently, the Iowa Electronic Markets have performed well in predicting the margin of victory for the presidential election winners.

Despite numerous attempts since the 1930s, no methods have been found to be superior to markets when forecasting prices. However, few people believe this as they pay handsomely for investment recommendations, a finding that has been labelled the 'Seer-sucker theory.' The theory states that no matter how much evidence there is that seers do not exist, suckers will continue to pay for the existence of seers (Armstrong 1980).

Software is available for creating trading platforms that allows participants to buy and sell contracts that represent their bets on events. Markets can be used to predict such things as the percentage of U.S. households with three or more vehicles by the end of 2015. Confidential betting markets can be set up within firms to bet on such things as first year sales of a new product. Some unpublished studies suggest that they can produce accurate sales forecasts for companies. However, there are no empirical studies that compare prediction market forecasts with those from traditional groups or from other methods.

Delphi

The Delphi technique was developed at RAND Corporation in the 1950s to help capture the knowledge of diverse experts while avoiding the disadvantages of traditional group meetings. The latter include group pressure, high administrative expenses, and lack of structure.

To forecast with Delphi, the administrator should ask between five and twenty experts who are diverse in their knowledge and opinions for their forecasts and their reasons for them. The administrator then provides the experts with anonymous summary statistics on the forecasts, and the experts' reasons for their forecasts. The process is repeated until there is little change in forecasts between rounds—two or three rounds are usually sufficient. The Delphi forecast is the median or mode of the experts' final-round forecasts. Software to help administer the procedure is available at forecastingprinciples.com.

Delphi has led to improved accuracy compared to forecasts from traditional groups in five studies, harmed accuracy in one, and was inconclusive in two in a meta-analysis by Rowe and Wright (2001). They found Delphi to be more accurate than one-round expert surveys for 12 of 16 studies, with two ties and two cases in which Delphi was less accurate. Over all of these 24 comparisons, Delphi improved accuracy in 71% and harmed it in 12%. On the other hand, Woudenberg's earlier (1991) review did not find an improvement in accuracy from Delphi. Delphi is likely to be most effective in situations where the relevant knowledge is distributed among the experts, such as in decisions on where to locate a retail outlet that would benefit from forecasts obtained from real estate, traffic, retailing, and consumer experts.

Delphi is attractive to managers because it is easy to understand and supports forecasts with reasons and authority (Green, Armstrong, & Graefe 2007). It is relatively cheap to conduct: panelists do not meet so the costs of assembling a group of highly-paid individuals in one place and the time-wasting of holding meetings are avoided. Moreover it is not necessary to employ expensive consultants to implement the method if a competent administrator can be found in-house.

Green, et al. (2007) identified eight advantage of the Delphi technique over prediction markets. These are in sum (1) broader applicability, (2) ease of understanding, (3) ability to address complex questions, (4) ability to maintain confidentiality, (5) avoidance of manipulation, (6) revelation of new knowledge, (7) avoidance of cascades, and (8) fewer participants. Points 6 and 7 refer to the fact that where the Delphi process requires participants to reveal their knowledge and reasoning and to respond to that of others, there is no such requirement on prediction market participants. As a consequence, prediction market participants might trade erroneously thinking that they have new information i.e. cascade.

Structured analogies

People often use analogies to make forecasts, but they tend to do so in an ad hoc manner. For example, they might search for a single analogy that suits their prior beliefs. The structured-analogies method uses a formal process to overcome biased and inefficient use of information from analogous situations.

To use the structured analogies method, an administrator prepares a description of the target situation and selects experts who are likely to know analogous situations; preferably from direct experience. The experts identify and describe analogous situations, rate their similarity to the target situation, and match the outcomes of their analogies with potential outcomes of the target situation. The administrator derives a forecast from each expert's analysis of his most similar analogy; one forecast per expert.

The limited research to date on structured analogies has been promising. Green and Armstrong (2007) found that structured analogies were 41% more accurate than unaided judgment in the difficult task of forecasting decisions in eight real conflicts, which included union-management disputes, a hostile takeover initiative, and a supply channel negotiation. The structured analogies method is especially appropriate when demand is affected by the actions of competitors, governments, or interest groups such as environmental and animal activists.

Game theory

The authors of textbooks and research papers recommend game theory for making forecasts about negotiations and other conflicts. Game theory involves identifying the incentives that motivate parties and deducing the decisions they will make in response to the incentives. On the face of it, this method might be useful for, for example, forecasting how competitors will react to a change in prices.

Despite the thousands of books on game theory, two papers by Green (2002 and 2005) provide the only evidence on the accuracy of game theorists forecasts. In these studies, game theory experts were urged to use game theory to predict the decisions that would be made in eight real conflict situations involving interaction. The game theorists' forecasts were no more accurate than those of university students using their unaided judgment.

Judgmental Decomposition

Judgemental decomposition involves dividing the forecasting problem into parts for which it is easier to derive forecast than it is for the whole problem. Different methods can be used for forecasting each part, as is appropriate to the nature of the part and the data available. The forecasts for the parts are combined to obtain a forecast for the whole.

One approach to decomposition is to break the problem down into multiplicative components. For example, to forecast sales for a brand, one might forecast market sales and market share separately, and then multiply the components. Forecasts from decomposition are generally more accurate than those from a global approach. In particular, decomposition is more accurate where there is much uncertainty about the aggregate forecast and where large numbers (over one million) are involved. MacGregor (2001, Exhibit 2) summarized results from three studies involving 15 tests and found that judgmental decomposition led to a 42% reduction in error when there was high uncertainty about the situation. Webby et al. (2005) found that a when forecasters were given help to decompose a time series forecasting task, their forecasts were more accurate.

Judgmental bootstrapping

Judgmental bootstrapping is used to estimate a formal forecasting model from experts' subjective judgments. Experts are asked what information they use to make predictions about a class of situations. They are then asked to make predictions for diverse cases, which can be real or hypothetical. For example, they might forecast first year turnover for stores using information about product range, proximity of competing stores, visibility, and traffic flows. These forecasts are used to estimate the coefficients of a regression equation that relates the experts' forecasts to the information they used. The general proposition seems preposterous: it is that the model of the man will be more accurate than the man. The reason is that the model applies the man's rules more consistently.

Judgemental bootstrapping models are most useful for repetitive complex forecasting problems where data on the dependent variable are not available (e.g. demand for a proposed sports stadium) or where the available data do not vary sufficiently for the estimation of an econometric model. Once developed, judgmental bootstrapping models provide a low-cost procedure for making forecasts.

Goodwin et al. (2011) suggested that judgmental bootstrapping is less likely to improve accuracy when many potential cues are available and it is not clear which ones the experts are using, where experts have access to information that is not available for the model or knowledge that cannot be readily incorporated into a regression model, or where cues (variables) are autocorrelated.

A meta-analysis of the evidence on judgmental bootstrapping found that forecasts were more accurate than those from unaided judgment (the normal method for these situations) in 8 of the 11 comparisons, with two tests showing no difference, and one showing a small loss (Armstrong 2001a). The typical error reduction was about 6%. The one failure occurred when the experts relied heavily on an erroneous variable.

Expert systems

Expert systems are structured implementations of the forecasting rules used by experts. One way to discover experts' rules is to create protocols by recording the experts as they talk about what they are doing while the make forecasts. Empirical estimates of relationships from structured analyses such as econometric studies and experiments should be used when available. Expert opinions, conjoint analysis, and bootstrapping can also provide useful information on rules. An expert system should be simple, clear, and complete.

In their review, Collopy, Adya and Armstrong (2001) found that expert systems forecasts were more accurate than those from unaided judgement. This conclusion was, however, based on only a small number of studies, and the gains in accuracy were small. Given the high cost of developing and revising expert systems, we expect that other methods will be more appropriate for most situations.

Simulated interaction

Simulated interaction is a form of role-playing that can be used to forecast decisions by people who are interacting with others. It is especially useful when the situation involves conflict. For example, a manager might want to know how best to secure an exclusive distribution arrangement with a major supplier, or how a competitor would respond to a 25% price reduction.

To use simulated interaction, an administrator prepares a description of the situation, describes the main protagonists' roles, and provides a short list of possible decisions. If necessary, secrecy can be maintained by disguising the situation. Role players adopt a role then read about the situation. They then engage in feasibly realistic interactions with the other role players until they reach a decision. The simulations usually last between 30 and 60 minutes.

Green (2005) found that relative to the usual forecasting method (unaided expert judgment), simulated interaction reduced forecast errors by 57% for the eight situations tested.

Simulated interaction is most useful when little or no quantitative data are available, the situation to forecast is unique or unusual, and decision makers wish to predict the effects of different policies or strategies. Simulated interactions can be conducted inexpensively by using students to play the roles. For example, it was used to determine how supermarkets would respond to a plan designed to give credit based on shopper's purchases so they could save money on the purchase of home appliances (Armstrong 2001c), and to predict the decision a company board would make over whether or not to make a contentious and substantial investment in new technology (Green 2002).

If the simulated interaction method seems onerous, you might wonder whether just following the common advice to "put yourself in the other person's shoes" will help you to predict the decisions they will make. It will not. Our study (Green and Armstrong 2011) failed to find any benefit from adopting this approach, even in a structured way. It is too difficult to think through the interactions of parties with divergent roles in a complex situation. Active role-playing between parties is needed to represent such situations with sufficient realism to derive useful forecasts.

Intentions and expectations surveys, and experimentation

Intentions surveys ask people how they *intend* to behave in specified situations. The data collected can be used, for example, to predict how people would respond to major changes in the design or price of a good. A meta-analysis covering 47 comparisons with over 10,000 subjects found that there is a strong relationship between people's intentions and their behavior (Kim and Hunter 1993).

Surveys can also be used to ask people how they *expect* they would behave. Expectations differ from intentions because people know that unintended things happen. For example, if you were asked whether you intended to visit the dentist in the next six months you might say no. However, you realize that a problem might arise that would necessitate a visit, so your expectation would be that visiting the dentist in the next six months had a probability greater than zero. Morwitz (2001) summarised evidence on expectations surveys.

To forecast demand using a survey of potential consumers, the administrator should prepare an accurate and comprehensive description of the product and conditions of sale. Expectations and intentions can be obtained using probability scales such as 0 = 'No chance, or almost no chance (1 in 100)' to 10 = 'Certain, or practically certain (99 in 100)'. Dillman (2000) provides evidence-based procedures for selecting samples, obtaining high response rates, compensating for non-response bias, and reducing response error. Response error (where respondent information is not accurately reported) is often a large component of error. This is especially so when the situation is new to people responding to the survey, as would be the case for questions about a new product. Wright and MacRae's (2007) meta-analysis concluded that intentions data provide unbiased forecasts of demand, so no adjustment is needed. Nevertheless, the expected errors are substantial on average.

Intentions and expectations surveys are especially useful when demand data are not available, such as for new product forecasts or for new markets for a product. They are most likely to be useful in cases where survey respondents have had relevant experience. Other conditions favouring the use of expectations surveys are: (1) responses can be obtained, (2) the behaviour is important to the respondent, (3) the behaviour is planned, (4) the plan is reported correctly, (5) the respondent is able to fulfil the plan, and (6) the plan is unlikely to change (Morwitz 2001).

Even better than surveys, are experiments. Experimentation can be used to test a planned change on a small scale in order to obtain evidence that can be used to forecast how the change will work out when implemented in full. Experiments can also be used to test alternative plans, in order to obtain predictions of the relative success of alternative courses of action. Finally, experiments can be used to estimate relationships. The relationship estimates can then be used to derive forecasting rules or models.

Experiments can be conducted in the field or in the "laboratory". Laboratory experiments allow greater control, testing of conditions is easier, they are typically cheaper to conduct, and they avoid revealing sensitive information to competitors prior to full implementation of a plan. A field experiment might involve, for example, distributing a product via a previously unused distribution channel in one of the company's territories and monitoring sales in order to estimate the effect on total revenue. A lab experiment might involve testing consumers' relative preferences by presenting a product in different packaging, giving participants a budget, and recording their purchases in a mock-up retail environment.

Focus group surveys are popular, but violate important forecasting principles and should not, therefore, be used to predict decisions or behaviour. First, focus groups are seldom representative of the population of interest—a focus group is a small sample—focus groups typically include six to ten individuals, whereas samples for intentions or expectations surveys tend to include several hundred people. Second, the responses of participants are influenced by the presence and expressed opinions of others in the group. Third, in practice, questions for the participants are generally not well structured or well tested. And, fourth, it is difficult to avoid subjectivity and bias in summarising the responses of focus group participants. There is no evidence to show that focus groups provide useful forecasts.

Methods requiring quantitative data

Extrapolation

Extrapolation methods require only sufficient historical data on the variable to be forecast. Statistical extrapolations are cost effective when many forecasts are needed. For example, some firms need forecasts of demand for each of hundreds of inventory items. They are also useful when little is known about the factors affecting a variable to be forecast (Armstrong 2001b).

Perhaps the most widely used extrapolation method, with the possible exception of using last year's value, is exponential smoothing. Exponential smoothing is understandable inexpensive, and relatively accurate. It implements the forecasting principles that recent data should be weighted more heavily, 'smoothes' cyclical fluctuations to forecast the trend, and can damp the trend if uncertainty is high. Gardner (2006) reviewed exponential smoothing methods—exponential smoothing is a family of methods with user-determined parameters—and the evidence on conditions and accuracy.

When extrapolating, remove the effect of seasonal influences first if data are shorter than annual. Makridakis *et al*. (1984) found that seasonality adjustments led to substantial gains in accuracy in their large-scale (xx forecasts for 1,428?? series) study of time-series forecasting. Estimates of seasonal factors should, however, be damped because seasonal adjustment programs tend to over-estimate seasonality.

Miller and Williams (2003, 2004) developed a procedure for damping seasonal factors estimated from historical data. When they applied the procedure to the 1,428 monthly time series from the M3-

Competition, their forecasts improved accuracy for 68% of the series. The Miller and Williams procedure is consistent with the principle that forecasters should be conservative in the face of uncertainty. Software for calculating damped seasonal adjustment factors is available at forecastingprinciples.com.

Sometimes managers know about external factors—such as a financial crisis in a major market—that will influence the variable being forecast. Rule-based forecasting, discussed below, integrates management knowledge with extrapolation forecasts.

Quantitative analogies

When few data are available on the thing being forecast, the target, quantitative data from analogous situations can be used to extrapolate what will happen. For example, in order to assess the annual percentage loss in sales when the patent protection for a drug is removed, one might examine the historical pattern of sales when patents were removed for similar drugs in similar markets.

To forecast using quantitative analogies, ask experts to identify situations that are analogous to the target situation and for which data are available. If the analogous data provide information about the target situation, such as per capita ticket sales for a play that is touring from city to city, forecast by calculating trimmed means.

Duncan, Gorr and Szczypula (2001) provided evidence that accuracy can be improved by using data from analogous time series. Averaging seasonal factors estimated from related series reduced forecast error by about 20% (Bunn and Vassilopoulos 1999) and pooling of seasonal crime rate factors across six precincts increased forecast accuracy by 7% (Gorr, Olligschlaeger and Thompson 2003). Thus, even when relevant data are available on the target, one should consider using analogous data to damp model coefficients or extrapolations towards the average for the class of situations.

Rule-based forecasting

Rule-based forecasting, or RBF, allows an analyst to integrate managers' knowledge about the situation with time-series data in a structured and inexpensive way. For example, in many cases, a useful guideline is that trends should be extrapolated only when they agree with managers' prior expectations. Forecast errors tend to be large when causal forces are contrary to the historical trend (Armstrong and Collopy 1993).

To use RBF, one must first identify features of the series. There are many features that are relevant to the selection of a method. A summary of 28 features is presented in Armstrong, Adya and Collopy (2001). They include forecast horizon, start-up series, bounded variables, number of observations, seasonality, and outliers. Features can be identified by inspection, statistical analysis, or domain knowledge (including causal forces). The RBF rules, there are 99 of them, are then used to adjust data, and to estimate short- and long-range models. RBF forecasts are a blend of the short- and long-range model forecasts.

RBF is most useful when substantive domain knowledge is available, patterns are discernable in the series, trends are strong, and forecasts are needed for horizons of six years or more. Under such conditions, rule-based forecast errors are substantially less than those for combinations of forecasts from other methods (Armstrong, Adya and Collopy 2001). In cases where the conditions were not met, the forecasts were no more accurate.

If implementing RBF is too big a step, firms should at least use the contrary series rule: When the expected and historical trends are "contrary" to one another, set the forecast trend to zero (Armstrong and Collopy 1993).

Neural nets

Neural nets are designed to pick up nonlinear patterns from long time-series. They have been of great interest to researchers. Wong, Lai and Lam (2000) found over 300 research papers published on neural nets during 1994 to 1998. Early reviews on the accuracy of forecasts from neural nets were not favorable. However, Adya and Collopy (1998) found eleven studies that met the criteria for a comparative evaluation, and in eight of these (73%), neural net forecasts were more accurate. While this is encouraging, our advice is to avoid neural networks because the method ignores prior knowledge and because the results are difficult to understand. . Furthermore, given the large number of studies on neural nets, the published research might not reflect the value of the method as studies with poor results might have been rejected due by journals due to the well-established bias against insignificant results. Perhaps the most extensive comparison was in the large-scale M3-Competition (Makridakis and Hibon 2000) with its 3,003 varied time series. In that study, neural net forecasts were 3.4% less accurate than damped trend forecasts and 4.2% less accurate than combined forecasts. Crone et al.'s (2011) NN3 competition, which was modeled on the M3 Competition and involved tests of *ex ante* accuracy in forecasting 111 or 11 empirical time series, found that neural network forecasts were comparably accurate to forecasts from established extrapolation methods, but not more accurate.

Causal Models

Causal models include models derived using regression analysis, the index method, and segmentation. These methods are useful if knowledge and data are available on variables that might affect the situation of interest. Forecasts from causal models were more accurate than forecasts derived from extrapolating the dependent variable when forecasting large changes (Armstrong 1985, p. 408-9; Allen and Fildes 2001). Theory, prior research, and expert domain knowledge provide information about relationships between the variable to be forecast and explanatory variables. When it is possible to include planning and decision-making variables in causal models, the models can be used to forecast the effects of different policies.

Causal models are most useful when (1) strong causal relationships exist, (2) the directions of the relationships are known, (3) there are large differences between alternatives (e.g. rival political candidates) or large changes are expected to occur in the causal variables over the forecast horizon, and (4) differences between alternatives are known or causal variables can be accurately forecast or controlled, especially with respect to their direction.

Regression analysis, or econometrics, involves estimating the coefficients of a causal model from historical data. These models estimate the relationship between a dependent variable and one or more explanatory or independent variables. They may be useful in situations in which three or fewer causal variables are important, effect sizes are important, effect sizes can be estimated from many reliable observations that include data in which the causal variables varied independently of one another, and the independent variables can be controlled or forecast accurately.

Important principles for developing regression models are to (1) use prior knowledge and theory, not statistical fit, for selecting variables and for specifying the directions of effects, (2) use simple models, (3) discard variables if the relationship estimated from the data conflicts with prior evidence on the nature of the relationship, and (4) keep the model simple in terms of the number of equations, number of variables and the functional form (Armstrong 1985).

Research findings on the importance of how well a model fits historical data are surprising. Model fit (often measured as R^2 or the standard error of the estimate) has a weak relationship with forecast accuracy and should be avoided. Instead, holdout data should be used as the benchmark against which to assess the predictive validity of a model. This conclusion is based on findings from many studies with time-series data (Armstrong, 2001e). Statistical fit has only a small relationship to forecast accuracy for cross-sectional data.

Because regression models tend to over-fit data, damping the estimated coefficients of a model tends to improve out-of-sample forecast accuracy, particularly if uncertainty is high, as occurs when one has small samples and many variables. As this situation is common for many prediction problems, unit (or equal weight) models – the most extreme case of damping – often yield more accurate forecasts than models with statistically fitted regression coefficients (Dana and Dawes, 2004). For example, Cuzán and

Bundrick (2009) found that equal-weight versions of prominent presidential election forecasting regression models provided more accurate forecasts than the original fitted models.

The *index method* is suitable for situations with little data on the variable to be forecast, where many causal variables are important, and where there is good prior knowledge about the effects of the variables. Use prior empirical evidence to identify predictor variables and to assess each variable's directional influence on the outcome. Experimental findings are especially valuable. Better yet, draw on findings from meta-analyses of experimental studies. If prior studies are not available, independent expert judgments can be used to choose the variables and determine the directions of their effects. If prior knowledge on a variable's effect is ambiguous or contradictory, do not include it in the model.

Index scores are the sum of the values across the variables, which might be coded as 0 or 1, or using a scale with more points, depending on the nature of the data. An alternative with the higher index score is "better". Where sufficient historical data are available, it is possible to estimate a forecasting model by regressing the index against the variable of interest, such as sales,

The index method is especially useful for selection problems, as for forecasting which advertisement will have the biggest effect on demand. Armstrong and Graefe (2011) describe the use of the method to make early forecasts of the outcomes of U.S. presidential elections from biographical information about potential candidates. Based on a list of 59 variables, the candidate's relative index scores correctly predicted the popular vote winner for 27 of the 29 elections from 1896 to 2008.

In cases where a single variable is more important than the rest of the variables, an accurate forecast can be made from the most important variable. This was used for example in the "take the best" model used to predict the outcome of political elections (Graefe and Armstrong 2010).

In general with causal models, one should avoid methods that lack theory or do not build upon prior knowledge. Rather than starting with theory, for example, data mining uses sophisticated statistical analyses to identify variables and relationships. Although it is a popular approach, we found no evidence that data-mining techniques provide benefits for forecasting. In their extensive review and their reanalysis of 50 real-world data sets, Keogh and Kasetty (2002) also found little evidence that data mining is useful. Much of the reason for the lack of evidence, they wrote, was the fact that few studies have used a proper design to assess data mining. This conclusion applies also to step-wise regression and neural networks.

Segmentation

Segmentation involves breaking a problem down into independent parts, using knowledge and data to make a forecast about each part, and combining the forecasts of the parts. For example, a hardware company could forecast sales for each type of product and then add the forecasts.

To forecast using segmentation, one must first identify important causal variables that can be used to define the segments, and their priorities (e.g. geographical location might have a higher priority than household income when forecasting demand for swimming pool chemicals). For each variable, cut-points are determined such that the stronger the relationship with the dependent variable, the greater the non-linearity in the relationship, and the more data that are available, the more cut-points should be used. Using the best method given the information available, forecasts are made for the population of each segment and the behaviour of the population within each segment. Population and behaviour forecasts are combined for each segment and the segment forecasts are summed.

Where there is interaction between variables, the effect of variables on demand are non-linear, and there are clear causal priorities, segmentation has advantages over regression analysis (Armstrong 1985, pp. 412-420). Segmentation is most useful when there are benefits from compensating errors. This is likely to occur where the segments are independent and are of roughly equal importance, and when information on each segment is good. For example, one might improve accuracy by forecasting demand for the products of each division of a company separately, then adding the forecasts. But if the segments are based on small samples and erratic data, the segment forecasts might contain large errors.

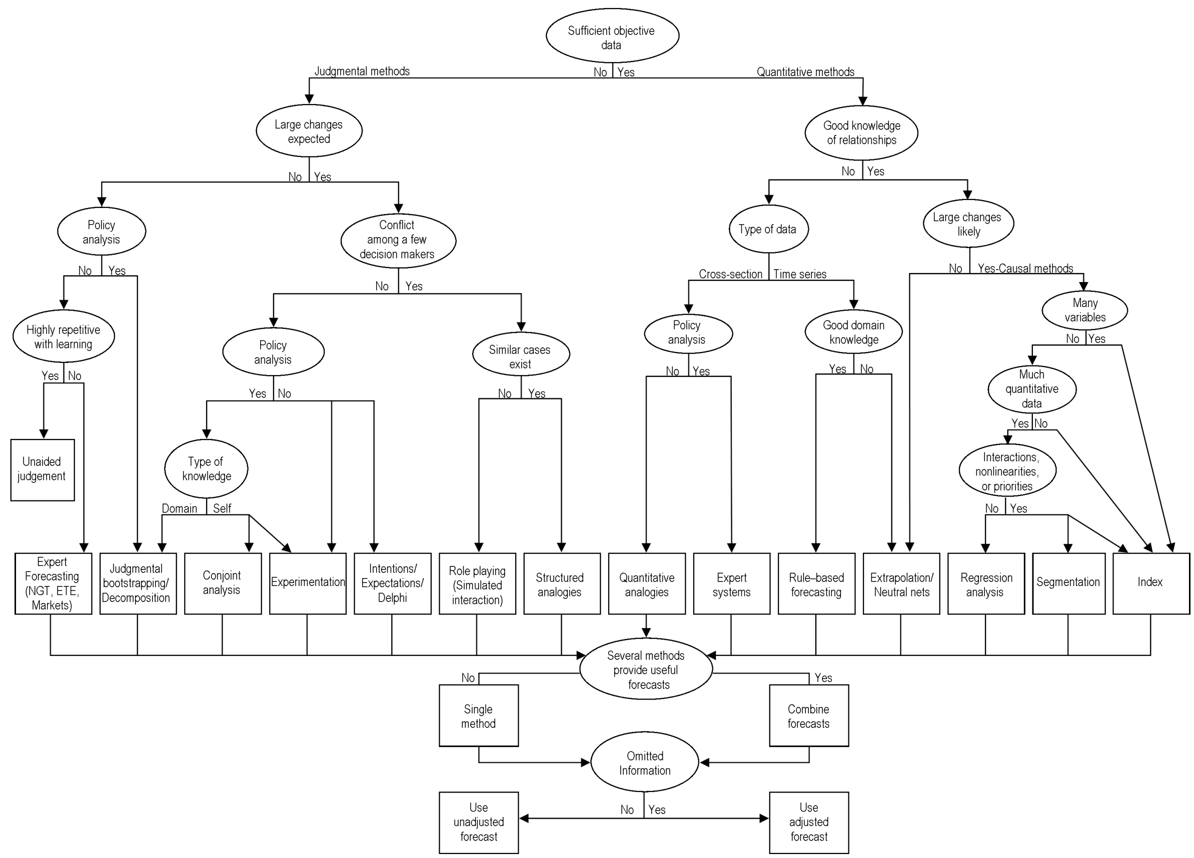
Segmentation based on *a priori* selection of variables does offer the possibility of improved accuracy at a low risk. Experts prefer the bottom-up approach of segmentation as it allows them to use their knowledge about the problem effectively (Jørgensen 2004). Dangerfield and Morris (1992), for example, found that bottom-up forecasting produced forecasts that were more accurate than those from top-down forecasting for 74% of the 192 monthly time series tested. In a study involving seven teams making estimates of the time required to complete two software projects, Jørgensen (2004) found that the typcal error from the bottom-up forecast was 51% less than that for the top-down approach.

In more recent studies, Carter and Chitturi (2009) compared the accuracy of forecasts of the diffusion of four new pharmaceutical products from three segmentation models with that of forecasts from a model that did not use segmentation. Errors of forecasts of 12 months of holdout data from the three segmentation models were, on average, 6.3%, 32%, and 61% smaller. Chang and Liao (2009) estimated models of Taiwan monthly outbound tourism numbers and for Hong Kong, Japan, and the U.S.A, using monthly data from 1996 to 2005. The errors for the individual country forecasts for 2006 averaged 77% of the MAPE for the total outbound forecasts. Chen *et al*. (2009) compared segmentation methods with no segmentation for forecasting housing prices in Knox County, Tennessee. Errors for forecasts from *a priori* segmentations were on average 7% smaller than for forecasts from no segmentation.

Selecting methods

Selecting the best forecasting method for a given situation is not a simple task. Often more than one will provide useful forecasts. In order to help forecasters choose methods that are appropriate for their problems, we used empirical findings and expert opinions to develop the decision tree shown in Figure 2.

Figure 2 Forecasting Method Selection Tree [Print landscape, full page]



The first question a forecaster confronts is whether the data are sufficient to develop quantitative models. If not, judgmental procedures are called for. Some situations call for both approaches.

For situations involving small changes, and where no policy analysis is needed and where one gets good feedback—such as with the number of diners that will come to a restaurant at a given time—unaided judgement can work well. If, however, the feedback is poor or uncertainty is high, it will help to use experts in a structured manner such as with a questionnaire or a prediction market or, if the relevant information is distributed among experts, with a Delphi panel. Where policy analysis is needed, judgemental bootstrapping or decomposition will help to use experts' knowledge effectively.

For situations involving large changes, but which do *not* involve conflicts among a few decision makers, ask next whether policy analysis is required. If policy analysis *is* required, judgmental bootstrapping and decomposition are structured methods for eliciting useful forecasts from experts. If people who are or might be involved in the situation being forecasts, as for instance potential customers, have knowledge about themselves that are relevant to predicting how people in the situation will behave, consider using conjoint analysis and experimentation. If policy analysis is not required, intentions or expectations surveys of, for example, potential customers may be useful, and the Delphi technique should also be considered. Experimentation is also likely to be useful.

To make forecasts about situations that involve conflict among a few decision makers, ask first whether similar cases exist. If they do, use structured analogies. If similar cases are hard to identify or the value of an accurate forecast is high, such as where a competitor reaction might have major consequences, use the simulated interaction method.

Turning now to situations where there are sufficient quantitative data to consider the estimation of quantitative models, ask first whether there is good knowledge about the relationships between causes and effects. If knowledge about such relationships is poor, speculative, or contentious, then consider next the kind of data that are available.

If the data are cross-sectional (e.g. for stores in different locations or product launches in different countries) use the method of quantitative analogies. For example, the introduction of new products in U.S. markets can provide analogies for the outcomes of the subsequent release of similar products in other countries. If analysis of alternative policies is needed, expert systems can be used. Expert systems are, however, expensive to develop.

If time-series data are available, use extrapolation methods to forecast. Where domain knowledge exists (such as when a manager knows that sales will increase due to the advertising of a price reduction), consider using rule-based forecasting. Much of the benefit of rule-based forecasting can be obtained by using the contrary series rule. The rule is easy to implement: ignore the historical trend when managers expect causal forces to act against the trend. For example, where sales of hats have been increasing over recent times, forecast flat sales when large net job losses are expected.

For situations where there is a good knowledge of relationships, extrapolation is appropriate if only small changes are expected, as is common in the short-term. If large changes are likely, causal methods are likely to provide forecasts that are more accurate than forecasts from other methods. Models estimated using regression analysis, or econometrics, may provide useful forecasts when there are few variables, much quantitative data, linear relationships, and an absence of interactions and priorities.

If the relationships are complicated, consider segmentation. Forecast the segments independently using appropriate methods.

Often the rigorous requisites of regression analysis are not met. In such situations, consider using the index method.

Combining and adjusting forecasts

Forecasters may have difficulty identifying which conditions apply to the situation they wish to forecast. In that case, use methods that draw on different assumptions about the condition, and then combine the forecasts according to rules determined before the forecasts were made.

Imagine you have two forecasts of a quantity: how big would the error of the second forecast need to be in order for the error of the average of the two forecasts to be larger than the error of the first forecast alone? The answer is that the error of the second forecast would need to be bigger than the error of the first forecast if the error was in the same direction (sign) but more than three times the error of the first forecast if the error had the opposite sign. In other words, if two forecasts bracket the outcome, it is highly likely that the combined forecast will be more accurate than the individual forecasts.

In order to increase the likelihood that two forecasts bracket the true value, use methods and data that differ substantially. The extent and probability of error reduction through combining is higher the greater the differences in the methods and data that produced the component forecasts (Batchelor and Dua 1995). For example, when combining real GNP forecasts, combining the 5% of forecasts that were most similar in their methods reduced the error compared to the typical forecast by 11%. By comparison, combining the 5% of forecasts that were most diverse in their methods yielded an error reduction of 23%.

Trimmed averages or medians are a good starting point for combining forecasts. Differential weights should only be used if there is strong evidence about the relative accuracy of forecasts from the different methods.

A meta-analysis of 30 studies found that combined forecasts (in most cases, simple averages of different forecasts from a single type of method) yielded a 12% reduction in error compared to the average error of the components. The reductions of forecast error ranged from 3 to 24%. In addition, the combined forecasts were often more accurate than the most accurate component (Armstrong, 2001d). Chen et al. (2009) found more modest error reductions (1-2%) from combining housing price forecasts from six methods—five of them segmentation methods—by simple average and by a variation of the 'encompass-combining algorithm'. Studies since that meta-analysis suggest that under favorable conditions (i.e., when forecasts are made for an uncertain situation, and many forecasts are available from several reasonable methods and different data sources) combining reduced errors almost by half (Graefe et al., 2010). Combining forecasts is especially useful if the forecaster wants to avoid large errors and if there is uncertainty about which method will be most accurate.

Combining improves accuracy, even if there are only two forecasts and both are from the same person. Herzog and Hertwig (2009) showed that combining the estimates of people who were asked to provide a second estimate that took into account knowledge they had previously ignored or rejected reduced error by 4.1%. However, combining the first estimates of two people was more accurate, with error reduced by 7.1%.

Judgmental and statistical methods should be integrated. Armstrong and Collopy (1998) summarized research in this area. Integration is effective when judgments are collected in a systematic manner and then used as inputs to the quantitative models, rather than simply used as adjustments to the outputs. Fildes et al. (2009) examined more than 60,000 forecasts from four supply chain companies and found judgmental adjustments of statistical demand forecasts increased forecast accuracy for three of the companies. Avoiding small adjustments and damping optimistic adjustments would have further increased accuracy for these companies.

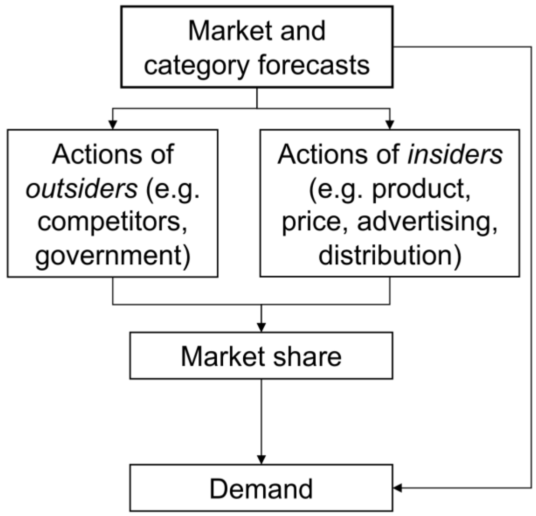
Goodwin (2005) provided evidence-based guidance on judgmental adjustments of statistical forecasts: (1) Adjust only for important information about future events; (2) Record reasons for adjustments; (3) Decompose the adjustment task if it is feasible to do so; (4) Mechanically combine judgmental and statistical forecasts; and (5) Consider using a Delphi panel for determining adjustments. Future events might include the implementation of a new government policy, a planned promotion, the loss of an important client, or a competitor's actions. Mechanical combination can be as simple as averaging. If the same forecaster makes many judgmental forecasts, the administrator or decision maker should, unbeknownst to the forecaster, consider estimating a regression model to correct the judgmental forecasts for biases. The process is known as Theil's correction (Goodwin et al. 2011).

When statistical forecasts are from causal methods, combining them with judgmental forecasts or adjustments can help accuracy if theory is inadequate to fully describe the process being forecast, important variables are missing from the causal model, data are poor, relationships are miss-specified, relationships are believed to have changed, or the environment has changed (Goodwin et al. 2011). Alternatively, regress judgmental forecast errors against the variables the forecasters could have been using and combine statistical forecasts of error from the resulting model with new judgmental forecasts to improve accuracy (Fildes's method; Fildes *et al* 2009).

On the need for forecasts

When one needs to understand why demand might change, or when changes are expected to be substantial, marketing managers may need to forecast the actions and reactions of key decision makers such as competitors, suppliers, distributors, collaborators, governments, and themselves. These actions can help to forecast market share. The resulting forecasts allow one to calculate a demand forecast as illustrated in Figure 3.

Figure 3
Needs for marketing forecasts



We now examine the need for the elements shown in Figure 3.

Forecasting market size

Market size is influenced by environmental factors such as economic conditions. For example, the demand for alcoholic beverages will be influenced by such things as local climate, size and age distribution of the population, distribution of disposable income, laws, culture, and religious beliefs. To forecast market size, one can use Delphi, intentions or expectations, extrapolation, causal methods, and segmentation.

Market forecasts for relatively new or rapidly changing markets in particular are often based on judgement. Given the risk of bias from unaided judgement, we recommend using structured methods. For example, the Delphi technique could be used to answer questions about market size such as: 'By what percentage will the wine market grow over the next 10 years?' or 'What proportion of households will watch movies via the Internet five years from now?"

When sufficient data are available, such as when the market is well established or when data on analogous markets or products are available, one can use either time-series extrapolation methods or causal methods. Simple time-series extrapolation is inexpensive. Causal methods, such as econometrics and segmentation, while more expensive, are likely to be the most accurate when large changes are expected in the causal variables, the direction of the change can be predicted accurately, and good knowledge exists about the effects of such changes.

Forecasting decision makers' actions

The development of a successful marketing strategy sometimes depends upon having good forecasts of the actions and reactions of competitors who might have an influence on market share. For example, if you lower your price, will competitors follow? A variety of judgmental methods can be used to forecast competitive actions. These include:

- expert opinion (ask experts who know about your and similar markets);
- intentions (ask competitors how they would respond in a given situation);
- structured analogies (analyse similar situations and the decisions that were made);

- simulated interaction (formal acting out of the interactions among decision makers for the firm and its competitors); and
- experimentation (trying the strategy on a small scale and monitoring the results).

Sometimes one may need to forecast the actions of other interest groups. For example, how would organizations that lobby for environmental causes react to the introduction of plastic packaging by a large fast-food restaurant chain? Structured analogies and simulated interaction would be useful here.

Company plans typically require the cooperation of many people. Managers may decide to implement a given marketing strategy, but will the organization be able to carry out the plan? Sometimes an organization fails to implement a plan because of a lack of resources, misunderstanding, or opposition by employees, or unions. The need to forecast organizational behaviour is sometimes overlooked. Better forecasting here might lead to more realistic plans and to plans that are easier to implement. Intentions surveys of key decision makers in an organization may help to assess whether a given strategy can be implemented successfully. Simulated interactions can also provide useful forecasts in such situations.

It is also important to predict the *effects* of the various actions. One can make such forecasts by using expert judgment, judgmental bootstrapping, or econometric methods.

Forecasting market share

If one expects the same causal forces and the same types of actions to persist into the future, a simple extrapolation of market share, such as from a naïve (e.g., constant market share) model, is usually sufficient.

When large changes are expected, one should draw upon methods that incorporate causal reasoning. If unusual changes are anticipated, judgmental methods such as Delphi would be appropriate. If the changes are expected to be large, the causes are well understood, and if data are scarce, judgmental bootstrapping can be used to improve forecasting.

The conditions for using econometric models for forecasting market share were described by Brodie, et al. (2001). Econometric methods should be used when (1) the effects of current marketing activity are strong relative to the residual effects of previous activity; (2) there are enough data and sufficient variability in the data; (3) models can allow for different responses by different brands; (4) models can be estimated at brand level; and (5) competitors actions can be forecast.

Knowledge about relationships can sometimes be can be obtained from prior research. For example, a meta-analysis by Tellis (1988) of price elasticities of demand for 367 branded products, estimated using econometric models, reported a mean value of -2.5. Hamilton, East, and Kilafatis (1997) analysis of 406 brand price elasticities also reported a value of -2.5. Estimates can also be made about other measures of market activity, such as advertising elasticity.

Forecasting for new products

The choice of a forecasting method depends on what stage it has reached in its life cycle. As a product moves from the concept phase to prototype, test market, introduction, growth, maturation, and declining stages, the relative value of the alternative forecasting methods changes. In general, the movement is from purely judgmental approaches to quantitative models.

New product forecasting is important given that large investments are commonly involved and uncertainty is high. This section considers methods that are relevant to new products.

Surveys of consumers' intentions and expectations are often used for new product forecasts. Intentions to purchase new products are complicated because potential customers may not be sufficiently familiar with the proposed product and because the various features of the product affect one another (e.g., price, quality and distribution channel). This suggests the need to prepare a good description of the circumstances surrounding the release of the proposed product, but a relatively simple description of the key features of the product may be sufficient (Armstrong and Overton 1971). A product description may involve prototypes, visual aids, product clinics or laboratory tests. Consumer surveys can also improve forecasts even when you already have some sales data (Armstrong, Morwitz, and Kumar 2000).

Errors in the description can be critical. For example, one of us was asked to forecast demand for the product of a new electricity retailer. As the retailer described the proposed product, an important feature was the ease with which customers would be able to swap their account to the new supplier. All they would have to do was to call the toll-free number and tell the operator their telephone number. Despite our concern that this level of ease might not be achievable, we proceeded to forecast demand using the electricity company's description. In the event, the existing supplier refused to transfer accounts without onerous proof, so demand was lower than predicted. This suggests the need to prepare alternative descriptions to examine the effect of different assumptions.

Conjoint analysis is widely used to see how demand varies as critical features of a product are varied (Wittink and Bergestuen 2001). Its use to forecast new-product demand can be expensive because it requires large samples of potential customers who may be difficult to identify, and the questionnaires are not easy for them to complete. Potential customers are asked to make selections from a set of offers such as 20 pairs of products. For example, various features of a personal digital assistant such as price, weight, battery life, and screen clarity could be varied substantially while ensuring that the variations in features do not correlate with one another. The potential customer choses from among various offerings in a way that is representative of how they would choose in the marketplace. The resulting data can be analysed by regressing respondents' choices against the product features.

The accuracy of forecasts from conjoint analysis is expected to increase with increasing realism of the choices presented to respondents). The method is based on sound principles, such as using experimental design and soliciting independent intentions from a representative sample of potential customers. Unfortunately however, there do not appear to be comparisons of conjoint-analysis forecasts with forecasts from other reasonable methods, despite repeated calls for such research (Wittink and Bergestuen 2001).

Expert opinions are widely used in the concept phase. For example, it is common to obtain forecasts from the sales force. It is important to properly pose the questions, adjust for biases in experts' forecasts, and aggregate their responses. The Delphi method provides an effective way to conduct such surveys.

Experts may be able to make better forecasts if the problem is decomposed in such a way that the parts to be forecast are better known to them than the whole. Thus, to forecast the sales of 3D television sets, rather than making a direct forecast, one could break the problem into parts such as 'How many households will there be in the U.S. in the forecast year?' 'Of these households, what percentage will make more than $30,000 per year?' and so on. The forecasts are obtained by multiplying the components.

Experts are often subject to biases when they make marketing forecasts. Those advocating a new product are likely to be optimistic. Sales people may try to forecast on the low side if their forecasts will be used to set quotas. Marketing executives may forecast high, believing that this will gain approval for a project or motivate the sales force. If possible, you should avoid experts who would have obvious reasons to be biased. Another strategy is to use a heterogeneous group of experts in the hope that their differing biases will tend to cancel one another.

Experts can make predictions about situations involving alternative product designs and alternative marketing plans. These predictions would then be related to the situations by regression analysis. It has advantages as compared to conjoint analysis in that few experts are needed (probably between five and twenty). In addition, expert judgments can incorporate policy variables, such as advertising, that are difficult for consumers to assess.

Analogous products can be used to forecast demand for new products. One collects a set of analogous products and examines their growth patterns as was done by Claycamp and Liddy (1969). The typical pattern can then be used as a forecast. While we are not aware of direct evidence on the accuracy of forecasts from this approach, evidence on structured analogies for forecasting decisions in conflicts (Green and Armstrong 2007), a difficult forecasting task, suggests that the forecasts would be useful.

Large errors are typical for new product forecasts. Tull (1967) estimated the mean absolute percentage error for new product sales to be about 65 percent.

Once a new product is on the market, it is possible to use extrapolation methods. For early sales, much attention has been given to the selection of the proper functional form. The diffusion literature uses an S-shaped curve to predict new product sales. That is, growth builds up slowly at first, becomes rapid if word-of-mouth is good, and if people see the product being used by others. Then it slows as it approaches a saturation level. A substantial literature exists on the most effective way to model this process. However, the number of comparative validation studies is small and the benefits of choosing the best functional form are modest (Meade and Islam 2001). Our advice is to use simple and understandable growth curves.

Uncertainty

In addition to improving accuracy, forecasting is concerned with assessing uncertainty. Good assessments of risk can help in planning, such as with the need for contingency plans.

Although the mean square error (MSE) appeals to statisticians, it does not provide a reliable basis for comparison of forecasting methods as shown in experiments by Armstrong and Collopy (1992). Judging from surveys described in McCarthy, et al. (2006), the use of the MSE has dropped substantially in recent years; their study estimated that, in recent years, only 6% of the firms use MSE. The median absolute percentage error (MdAPE) is more appropriate because it is not affected by scale or by outliers.

Tests of statistical significance should not be used for assessing uncertainty. As shown by Schmidt and Hunter (1997), statistical significance is of no value for analyzing data, even when properly used and properly interpreted. We have been unable to find a single case where statistical significance has made a contribution to forecasting (Armstrong 2007). For a comprehensive review of the evidence on the value of tests of statistical significance, see Ziliak & McCloskey (2008).

Instead of statistical significance, the focus should be on prediction intervals (confidence intervals). Chatfield (2001) summarizes research on prediction intervals. Unfortunately, prediction intervals are not widely used in practice. Dalrymple (1987) found that 48% of firms did not use confidence intervals, and only 10% 'usually' used them.

The fit of a model to historical data is a poor way to estimate prediction intervals. It typically results in intervals that are too narrow. It is best to simulate the actual forecasting procedure as closely as possible, and use the distribution of the resulting *ex ante* forecasts to assess uncertainty. For example, if you need to make forecasts for two years ahead, withhold enough data to be able to estimate the forecast error for two-year-ahead *ex ante* forecasts.

Experts are typically overconfident (Arkes, 2001). In an examination of economic forecasts from 22 economists over 11 years, the actual values fell outside the range of their prediction intervals about 43% of the time, and this occurred even when subjects were warned in advance against overconfidence (McNees 1992). To improve the calibration of judges, they should receive timely and unambiguous information on outcomes, along with reasons why they were right or wrong. When feedback is good, judges' confidence intervals are well calibrated. For example, for days when weather forecasters say that there is a 60% chance of rain, it rains on 60% of the days. This suggests that marketing forecasters would do well to seek accurate and well-summarized feedback. In cases where good feedback is not possible, ask experts to write all the reasons why their forecasts might be wrong (Arkes, 2001).

Still another way to assess uncertainty is to examine the agreement among judgmental forecasts. For example, Ashton (1985), in a study of forecasts of annual advertising sales for *Time* magazine, found that the agreement among the individual judgmental forecasts was a good proxy for uncertainty.

Prediction intervals from quantitative forecasts tend to be too narrow even when based on *ex ante* forecasts. Some empirical studies have shown that the percentage of actual values that fall outside the 95% prediction intervals is often greater than 50% (Makridakis, et al. 1987). This occurs because the estimates ignore some sources of uncertainty and because discontinuities sometimes occur over the forecast horizon. In addition, forecast errors in time series are often asymmetric, so this makes it difficult to estimate prediction intervals. Asymmetry of errors is likely to occur when the forecasting model uses an additive trend. The most sensible procedure is to transform the forecast and actual values to logs, then calculate the prediction intervals using logged differences (but presenting the findings in actual values). Interestingly, researchers and practitioners seldom follow this advice. Evidence on the issue of asymmetrical errors is provided in Armstrong and Collopy (2001).

Loss functions can also be asymmetric. For example, the losses due to a forecast that is too low by 50 units may differ from the losses if it is too high by 50 units. But this is a problem for the planner, not the forecaster.

Gaining acceptance of forecasts

Forecasts that contradict management's expectations can be valuable. Unfortunately they may also be ignored as was shown in a study by Griffith and Wellman (1979). One way to avoid this problem is to gain prior agreement from managers on the forecasting procedures to use. Another way to increase the likelihood that forecasts will be accepted is to ask decision makers to determine in advance what decisions they will make when presented with different possible forecasts. If the decisions would not be affected by the forecasts, there is no need to make forecasts.

The use of scenarios can help to gain acceptance of forecasts. Scenarios involve providing information about a postulated future situation to decision makers and asking them to project themselves into the situation and to write stories about how they got to that situation. The stories should be written in the past tense. More detailed instructions for writing scenarios are provided in Gregory and Duran (2001). Scenarios are effective in getting managers to accept the possibility that certain events might occur.

Scenarios should *not*, however, be used to make forecasts, because they inflate managers' assessment of the likelihood that the events they described will occur far beyond any objective probability.

Conclusions

Important advances have been made in forecasting over the past half century. These advances can be used to improve many aspects of forecasting demand. Some advances relate to judgment, such as Delphi, simulated interactions, intentions studies, opinions surveys, judgmental bootstrapping, and combining. Others relate to quantitative methods such as extrapolation, rule-based forecasting, and the index method. Most recently, gains have come from the integration of statistical and judgmental forecasts. Finally, we have learned much about how to gain acceptance of forecasts.

Most firms ignore or are unaware of the evidence-based techniques and principles for forecasting. Consequently, there are many opportunities for improvement in forecasting practice. This is partly because the evidence has been hidden in sometimes-turgid academic journal articles. Over the past few years, much effort has been devoted to improving the situation for practitioners by providing understandable principles that summarize these findings. These evidence-based principles are freely available at forecastingprinciples.com.

REFERENCES

Adya, Monica and Fred Collopy. 1998. "How effective are neural nets at forecasting and prediction? A review and evaluation." *Journal of Forecasting* 17: 451-461.

Allen, P. Geoffrey and Robert Fildes. 2001. "Econometric forecasting." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 303-362.

Arkes, Hal R. 2001. "Overconfidence in judgmental forecasting." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 495-515.

Armstrong, J. Scott 2007, "Significance tests harm progress in forecasting," *International Journal of Forecasting*, 23, 321-327
Armstrong, J. Scott. Ed. 2001. *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
Armstrong, J. Scott. 2001a. “Judgmental bootstrapping: Inferring experts’ rules for forecasting.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 171-192.
Armstrong, J. Scott. 2001b. “Extrapolation of time-series and cross-sectional data.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 217-243.
Armstrong, J. Scott. 2001c. “Role playing: A method to forecast decisions.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 13-30.
Armstrong, J. Scott. 2001d. “Combining forecasts.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 417-439.
Armstrong, J. Scott. 2001e. “Evaluating forecasting methods.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 365-382.
Armstrong, J. Scott, Monica Adya, and Fred Collopy. 2001. “Rule-based forecasting: Using judgment in time-series extrapolation.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 259-282.
Armstrong, J. Scott and Fred Collopy. 2001. “Identification of asymmetric prediction intervals through causal forces.” *Journal of Forecasting* 20: 273-283.
Armstrong, J. Scott and Fred Collopy. 1998. “Integration of statistical methods and judgment for time series forecasting: Principles from empirical research.” In *Forecasting with Judgment*. Eds. George Wright and Paul Goodwin. Chichester: John Wiley.
Armstrong, J. Scott and Fred Collopy. 1993. “Causal forces: Structuring knowledge for time series extrapolation.” *Journal of Forecasting* 12: 103-115.
Armstrong, J. Scott and Fred Collopy. 1992. “Error measures for generalizing about forecasting methods: empirical comparisons.” *International Journal of Forecasting* 8: 69-80.
Armstrong, J. S. and Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method, *Journal of Business Research* 64, 699-706,
Armstrong, J. Scott, Vicki Morwitz, and V. Kumar. 2000. “Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy?” *International Journal of Forecasting* 16: 383-397.
Armstrong, J. Scott & T. S. Overton (1971), “Brief vs. Comprehensive Descriptions in Measuring Intentions to Purchase,” *Journal of Marketing Research,* 8 (1971), 114-117.
Ashton, Alison H. 1985. “Does consensus imply accuracy in accounting studies of decision making?” *Accounting Review* 60: 173-185.
Batchelor, Roy and Pami Dua. 1995. “Forecaster diversity and the benefits of combining forecasts.” *Management Science* 41: 68-75.
Brodie, Roderick J., Peter Danaher, V. Kumar, and Peter Leeflang. 2001. “Econometric models for forecasting market share.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 597-611.
Carter, Franklin, J., and Ravindra Chitturi. 2009. “Segmentation Based on Physician Behavior: Implications for Sales Forecasting and Marketing-Mix Strategy,” *Journal of Personal Selling & Sales Management 29*: 81-95.
Chang, Yu-Wei, and Meng-Yuan Liao. 2010. “A seasonal ARIMA model of tourism forecasting: The case of Taiwan” *Asia Pacific Journal of Tourism Research* 15(2): 215-221.
Chatfield, Christopher. 2001. “Prediction intervals for time series.” In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 475-494.
Chen, Zhuo, Seong-Hoon Cho, Neelam Poudyal, and Roland K. Roberts. 2009. “Forecasting housing prices under different market segmentation assumptions” *Urban Studies* 46(1): 167-187.
Claycamp, Henry J. and Lucien E. Liddy. 1969. “Prediction of new product performance: An analytical approach.” *Journal of Marketing Research* 6: 414-420.

Collopy, Fred, Monica Adya, and J. Scott Armstrong. 2001, "Expert systems for forecasting." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 285-300

Crone, Sven F., Michèle Hibon, and Konstantinos Nikolopoulos. 2011, "Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction." *International Journal of Forecasting* 27: 635-660.

Dana, Jason and Robyn M. Dawes. 2005, "The superiority of simple alternatives to regression for social science predictions." *Journal of Educational and Behavioral Statistics* 29 (3): 317-331.

Dangerfield, Byron J. and John S. Morris. 1992. "Top-down or bottom-up: Aggregate versus disaggregate extrapolations." *International Journal of Forecasting* 8: 233-241.

Dillman, Don. A. 2000. *Mail and Internet Surveys: The Tailored Design Method, (2nd ed.).* New York: John Wiley.

Duncan, George T., Wilpen L. Gorr, and Janusz Szczypula. 2001, "Forecasting analogous time series." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 195-213.

Gardner, Everette S., Jr. 2006. "Exponential smoothing: The state of the art – Part II (with commentary)." *International Journal of Forecasting* 22: 637-677.

Fildes, R., P. Goodwin, M. Lawrence, and K. Nikolopoulos. 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* 25:3–23.

Goodwin, P. (2005). How to integrate management judgment with statistical forecasts. *Foresight*, *1*, 8-12.

Goodwin, P., Önkal, D., & Lawrence, M. (2011). Improving the role of judgment in economic forecasting. In Clements, M. P. & Hendry, D. F. (Eds.) *Oxford Handbook of Economic Forecasting*, OUP: Oxford, UK. 163–189.

Graefe, A., Armstrong, J. S., Jones, R. J. and Cuzán, A. G. (2010). Combining forecasts: An application to U.S. Presidential Elections, *Working paper*, Available at: http://dl.dropbox.com/u/3662406/Articles/Graefe_et_al_Combining.pdf

Graefe, A. & J. Scott Armstrong (2010), Predicting elections from the most important issue: A test of the take-the-best heuristic, *Journal of Behavioral Decision Making*, Published online at http://onlinelibrary.wiley.com/doi/10.1002/bdm.710/full

Graefe, A. & J. Scott Armstrong (2011), "Conditions under which index models are useful: Reply to Bio-index Commentaries" *Journal of Business Research*, *64*, 693-695.

Gregory, W. Larry and Anne Duran. 2001. "Scenarios and acceptance of forecasts." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 519-541.

Green, Kesten C. 2005. "Game theory, simulated interaction, and unaided judgment for forecasting decisions in conflicts: Further evidence." *International Journal of Forecasting* 21: 463-472.

Green, Kesten C. 2002. "Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement." *International Journal of Forecasting* 18: 321-344.

Green, Kesten C. and J. S. Armstrong, 2011. "Role Thinking: Standing in Other People's Shoes to Forecast Decisions in Conflicts," *International Journal of Forecasting*, 27, 69-80.

Green, Kesten C. and J. Scott Armstrong, 2007. "Structured analogies for forecasting." *International Journal of Forecasting* 23, 365-376

Green, K. C., Armstrong, J. S., & Graefe, A. 2007. Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *Foresight*, *8*, 17-20. Available from http://kestencgreen.com/green-armstrong-graefe-2007x.pdf

Griffith, John R. and Barry T. Wellman. 1979. "Forecasting bed needs and recommending facilities plans for community hospitals: A review of past performance." *Medical Care* 17: 293-303.

Hamilton, Will, Robert East, and Stavros Kilafatis. 1997. "The measurement and utility of brand price elasticities." *Journal of Marketing Management* 13: 285-298.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*, 231-237.

Jørgensen, Magne. 2004, "Top-down and bottom-up expert estimation of software development effort." *Journal of Information and Software Technology* 46 (1): 3-16.

Juster, Thomas. 1966. "Consumer buying intentions and purchase probability: An experiment in survey design." *Journal of the American Statistical Association* 61: 658-696.
Keogh, Eamonn J. and Shruti Kasetty. 2002. "On the need for time series data mining benchmarks: A survey and empirical demonstration." Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Kim, M. & Hunter, J. E. (1993). Relationships among attitudes, behavioral intentions, and behavior: A meta-analysis of past research, *Communication Research*, 20, 331-364.
MacGregor, Donald G. 2001. "Decomposition for judgmental forecasting and estimation." in J. S. Armstrong (Ed.) *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers, 107-123.
Makridakis, Spyros G., A. Andersen., R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. 1984. *The Forecasting Accuracy of Major Times Series Methods.* Chichester: John Wiley.
Makridakis, Spyros G., Michèle Hibon, F. Lusk, and M. Belhadjali. 1987. "Confidence intervals: An empirical investigation of tine series in the M-competition." *International Journal of Forecasting* 3: 489-508.
Makridakis, Spyros G. and Michèle Hibon. 2000, "The M3-Competition: Results, conclusions and implications." *International Journal of Forecasting* 16: 451-476.
Makridakis, Spyros G., Steven C. Wheelwright, and Rob J. Hyndman. 1998. *Forecasting Methods for Management,* Third edition. New York: John Wiley.
McCarthy, Teresa M., Donna F. Davis, Susan L. Golicic, and John T. Mentzer. 2006. "The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices." *Journal of Forecasting* 25: 303-324.
McNees, Stephen K. 1992. "The uses and abuses of 'consensus' forecasts." *Journal of Forecasting* 11: 703-710.
Meade, Nigel and Towhidul Islam. 2001. "Forecasting the diffusion of innovations: Implications for time series extrapolation." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 577-595.
Miller, Don M. and Dan Williams. 2004. "Shrinkage estimators for damping X12-ARIMA seasonals." *International Journal of Forecasting* 20: 529-549.
Morwitz, Vicki G. 2001. "Methods for forecasting from intentions data." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 33-56.
Rhode, Paul. W. and Koleman S. Strumpf. 2004. "Historical presidential betting markets." *Journal of Economic Perspectives* 18 (2): 127-142.
Rowe, Gene and George Wright. 2001. "Expert opinions in forecasting role of the Delphi technique." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 125-144.
Schmidt, Frank L. and John E. Hunter. 1997. "Eight common but false objections to the discontinuation of significance testing in the analysis of research data." In *What if there were no Significance Tests?* Lisa L. Harlow, Stanley A. Mulaik, and James H. Steiger, Eds London: Lawrence Erlbaum, 37-64.
Tellis, Gerald J. 1988. "The price elasticity of selective demand: A meta-analysis of econometric models of sales." *Journal of Marketing Research* 25: 331-341.
Tellis, Gerald J. 2009. "Generalizations about advertising effectiveness in markets," *Journal of Advertising Research*, 49,240-245.
Tetlock, Philip E. (2005). *Expert political judgment: How good is it? How can we know?* New Jersey: Princeton University Press.
Tull, Donald. S. 1967. "The relationship of actual and predicted sales and profits in new product introductions." *Journal of Business* 40: 233-250.
Webby, W., M. O'Connor, and B. Edmundson. 2005. Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting 21*:411–423.
Wittink, Dick R. and Trond. Bergestuen. 2001. "Forecasting with conjoint analysis." In *Principles of Forecasting*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers, 147-167.

Wright, Malcolm. & Murray MacRae 2007, "Bias and variability in purchase intention scales," *Journal of the Academy of Marketing Science*, 35, 617- 624.

Wong, Bo K, Vincent S. Lai, and Jolie Lam. 2000. "A bibliography of neural network business applications research: 1994-1998." *Computers & Operations Research* 27: 1045-1076.

Woudenberg, F. (1991). An Evaluation of Delphi, *Technological Forecasting and Social Change*, 40, 131-150.

Ziliak, S. T. & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Authors

J. Scott Armstrong (Ph.D., MIT, 1968), Professor of Marketing at the Wharton School, University of Pennsylvania, is a founder of the *Journal of Forecasting*, the *International Journal of Forecasting*, and the International Symposium on Forecasting. He is the creator of forecastingprinciples.com and editor of *Principles of Forecasting* (Kluwer 2001), an evidence-based summary of knowledge on forecasting. In 1996, he was selected as one of the first "Honorary Fellows" by the International Institute of Forecasters. In 2004 and 2008, his PollyVote.com team showed how scientific forecasting principles can produce highly accurate forecasts of US presidential elections. He was named by the Society of Marketing Advances as "Distinguished Marketing Scholar of 2000." One of Wharton's most prolific scholars, he is the most highly cited professor in the Marketing Department at Wharton. His current projects involve the application of scientific forecasting methods to climate change, the effectiveness of learning at universities, and the use of the index method to make predictions for situations with many variables and much knowledge. His book, *Persuasive Advertising,* was published by Palgrave Macmillan in 2010. It summarizes evidence-based knowledge on persuasion and it is supported by advertisingprinciples.com. He can be contacted at Armstrong@wharton.upenn.edu.

Kesten C. Green (Ph.D., VUW, 2002) is a Senior Lecturer in the International Graduate School of Business of the University of South Australia and a Senior Research Associate of the Ehrenberg-Bass Institute for Marketing Science. He is also a Director of the International Institute of Directors and co-director of the Forecasting Principles public service Internet site devoted to the advancement of evidence-based forecasting. His research has led to improvements in forecasting the decisions people make in conflicts such as occur in business competition, supply chains, mergers and acquisitions, and between customers and businesses. His other interests include forecasting for public policy, forecasting demand, forecasting for recessions and recoveries, and the effect of business objectives on performance. His research has been covered in the *Australian Financial Review*, the *London Financial Times*, the *New Yorker*, and the *Wall Street Journal*. He has advised the Alaska Department of Natural Resources, the U.S. Department of Defense, the Defense Threat Reduction Agency, the National Security Agency (NSA) and more that 50 other business and government clients.