12-1-2010

Forecasting Principles

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

Kesten C. Green
University of South Australia

Andreas Graefe
Karlsruhe Institute of Technology

Suggested Citation:
Armstrong, J.S., Green, K.C. and Graefe, A. "Forecasting Principles." In *International Encyclopedia of Statistical Science* (Ed. M. Lovric). Springer, 2011.
http://dx.doi.org/10.1007/978-3-642-04898-2_257

Forecasting Principles

Working Paper

Published in M. Lovric, *International Encyclopedia on Statistical Science* (2010)

Kesten C. Green
International Graduate School of Business, University of South Australia
Adelaide, SA 5000, Australia
Kesten.Green@unisa.edu.au

Andreas Graefe
Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Germany
Graefe@kit.edu

J. Scott Armstrong
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
Armstrong@wharton.upenn.edu

Forecasting is concerned with making statements about the as yet unknown. There are many ways that people go about deriving forecasts. This entry is concerned primarily with procedures that have performed well in empirical studies that contrast the accuracy of alternative methods.

Evidence about forecasting procedures has been codified as condition-action statements, rules, guidelines or, as we refer to them, *principles*. At the time of writing there are 140 principles. Think of them as being like a safety checklist for a commercial airliner—if the forecast is important, it is important to check all relevant items on the list. Most of these principles were derived as generalized findings from empirical comparisons of alternative forecasting methods. Interestingly, the empirical evidence sometimes conflicts with common beliefs about how to forecast.

Primarily due to the strong emphasis placed on empirical comparisons of alternative methods, researchers have made many advances in forecasting since 1980. The most influential paper in this regard is the M-competition paper (Makridakis *et al.* 1982). This was based on a study where different forecasters were invited to use what they thought to be the best method to forecast many times series. Entry into the competition required that methods were fully disclosed. Entrants submitted their forecasts to an umpire who calculated the errors for each method. This was only one in a series of M-competition studies, the most recent being Makridakis and Hibon (2000). For a summary of the progress that has been made in forecasting since 1980, see Armstrong (2006).

We briefly describe valid forecasting methods, provide guidelines for the selection of methods, and present the *Forecasting Canon* of nine overarching principles. The *Forecasting Canon* provides a gentle introduction for those who do not need to become forecasting experts but who nevertheless rightly believe that proper knowledge about forecasting would help them to improve their decision making. Those who wish to know more can find what they seek in *Principles of Forecasting: A Handbook for Practitioners and Researchers*, and at the Principles of Forecasting Internet site (ForPrin.com).

Forecasting methods

As shown in Figure 1, the *Forecasting Methodology Tree*, forecasting methods can be classified into those that are based primarily on judgmental sources of information and those that use statistical data. There is overlap between some judgmental and statistical approaches.

—— Figure 1 (Methodology Tree) about here ——

If available data are inadequate for quantitative analysis or qualitative information is likely to increase the accuracy, relevance, or acceptability of forecasts, one way to make forecasts is to ask experts to think about a situation and predict what will happen. If experts' forecasts are not derived using structured forecasting methods, their forecasting method is referred to as *unaided judgment*. This is the most commonly used method. It is fast, inexpensive when few forecasts are needed, and may be appropriate when small changes are expected. It is most likely to be useful when the forecaster knows the situation well and gets good feedback about the accuracy of his forecasts (e.g., weather forecasting, betting on sports, and bidding in bridge games).

Expert forecasting refers to forecasts obtained in a structured way from two or more experts. The most appropriate method depends on the conditions (e.g., time constraints, dispersal of knowledge, access to experts, expert motivation, need for confidentiality). In general, diverse experts should be recruited, questions should be chosen carefully and tested, and procedures for combining across experts (e.g., the use of medians) should be specified in advance.

The *nominal group technique* (NGT) tries to account for some of the drawbacks of traditional meetings by imposing a structure on the interactions of the experts. This process consists of three steps: First, group members work independently and generate individual forecasts. The group then conducts an unstructured discussion to deliberate on the problem. Finally, group members work independently and provide their final individual forecasts. The NGT forecast is the mean or median of the final individual estimates.

Where group pressures are a concern or physical proximity is not feasible, the *Delphi method,* which involves at least two rounds of anonymous interaction, may be useful. Instead of direct interaction, individual forecasts and arguments are summarized and reported as feedback to participants after each round. Taking into account this information, participants provide a revised forecast for the next round. The Delphi forecast is the mean or median of the individual forecasts in the final round. Rowe and Wright (2001) found that Delphi improved accuracy over unstructured groups in five studies, harmed accuracy in one, and the comparison was inconclusive in two. Delphi is most suitable if experts are expected to possess different information, but it can be conducted as a simple one-round survey for situations in which experts possess similar information. A free version of the Delphi software is available at ForPrin.com.

In situations where dispersed information frequently becomes available, *prediction markets* can be useful for providing continuously updated numerical or probability forecasts. In a prediction market, mutually anonymous participants reveal information by trading contracts whose prices reflect the aggregated group opinion. Incentives to participate in a market may be monetary or non-monetary. Although prediction markets seem promising, to date there has been no published meta-analysis of the method's accuracy. For a discussion of the relative advantages of prediction markets and Delphi see Green *et al*. (2007).

With *structured analogies*, experts identify situations that are analogous to a target situation, identify similarities and differences to the target situation, and determine an overall similarity rating. The outcome or decision implied by each expert's top-rated analogy is used as the structured analogies forecast. Green and Armstrong (2007) analyzed structured analogies for the difficult problem of forecasting decisions people will make in conflict situations. When experts were able to identify two or more analogies and their closest analogy was from direct experience, 60% of structured analogies forecasts were accurate compared to 32% of experts' unaided judgment forecasts, the latter being little better than guessing.

Decomposition involves breaking down a forecasting problem into components that are easier to forecast. The components may either be multiplicative (e.g., to forecast a brand's sales, one could estimate total market sales and market share) or additive (estimates could be made for each type of product when forecasting new product sales for a division). Decomposition is most likely to be useful in situations involving high uncertainty, such as when predicting large numbers. MacGregor (2001) summarized results from three studies involving 15 tests and found that judgmental decomposition led to a 42% reduction in error under high levels of uncertainty.

Judgmental bootstrapping derives a model from knowledge of experts' forecasts and the information experts used to make their forecasts. This is typically done by regression analysis. It is useful when expert judgments have validity but data are scarce (e.g., forecasting new products) and outcomes are difficult to observe (e.g., predicting performance of executives). Once developed, judgmental bootstrapping models are a low-cost forecasting method. Armstrong (2001a) found judgmental bootstrapping to be more accurate than unaided judgment in 8 of 11 comparisons. Two tests found no difference, and one found a small loss in accuracy.

Expert systems are based on rules for forecasting that are derived from the reasoning experts use when making forecasts. They can be developed using knowledge from diverse sources such as surveys, interviews of experts, protocol analysis in which the expert explains what he is doing as he makes forecasts, and research papers. Collopy *et al*. (2001) summarized evidence from 15 comparisons that included expert systems on the predictive validity of the method. Expert systems were more accurate than unaided judgment in six comparisons, similar in one, and less accurate in another. Expert systems were less accurate than judgmental bootstrapping in two comparisons and similar in two. Expert systems were more accurate than econometric models in one comparison and as accurate in two.

It may be possible to ask people directly to predict how they would behave in various situations. However, this requires that people have valid *intentions* or *expectations* about how they would behave. Both are most useful when (1) responses can be obtained from a representative sample, (2) responses are based on good knowledge, (3) people have no reason to lie, and (4) new information is unlikely to change behavior. Intentions are more limited than expectations in that they are most useful when (5) the event is important, (6) the behavior is planned, and (7) the respondent can fulfill the plan (e.g., their behavior is not dependent on the agreement of others).

Role playing involves asking people to think and behave in ways that are consistent with a role and situation described to them. Role playing for the purpose of predicting the behavior of people with different roles who are interacting with each other is called *simulated* *interaction*. Role players are assigned roles and asked to act out prospective interactions in a realistic manner. The decisions are used as forecasts of the actual decision. Green (2005) found that 62% of simulated interaction forecasts were accurate for eight diverse conflict situations. By comparison, 31% of forecasts from the traditional approach—expert judgments unaided by structured techniques—were accurate. Game theory experts' forecasts were no better, also 31%, and both unaided judgment and game theory forecasts were little better than chance at 28% accurate.

Conjoint analysis is a method for eliciting people's preferences for different possible offerings (e.g. for alternative mobile phone designs or for different political platforms) by using combinations of features (e.g. size, camera, and screen of a mobile phone.) The possibilities can be set up as experiments where each variable is unrelated to the other variable. Regression-like analyses are then used to predict the most desirable design.

Extrapolation models use time-series data on the situation of interest (e.g., data on automobile sales from 1940-2009) or relevant cross-sectional data. For example, exponential smoothing, which relies on the principle that more recent data is weighted more heavily, can be used to extrapolate over time. Quantitative extrapolation methods do not harness people's knowledge about the data but assume that the causal forces that have shaped history will continue. If this assumption turns out to be wrong, forecast errors can be large. As a consequence, one should only extrapolate trends when they correspond to the prior expectations of domain experts. Armstrong (2001b) provides guidance on the use of extrapolation.

Quantitative analogies are similar to structured analogies. Experts identify analogous situations for which time-series or cross-sectional data are available, and rate the similarity of each analogy to the data-poor target situation. These inputs are used to derive a forecast. This method is useful in situations with little historical data. For example, one could average data from cinemas in suburbs identified by experts as similar to a new (target) suburb in order to forecast demand for cinema seats in the target suburb.

Rule-based forecasting is an expert system for combining expert domain knowledge and statistical techniques for extrapolating time series. Most series features can be identified automatically, but experts are needed to identify some features, particularly causal forces acting on trends. Collopy and Armstrong (1992) found rule-based forecasting to be more accurate than extrapolation methods.

If data are available on variables that might affect the situation of interest, causal models are possible. Theory, prior research, and expert domain knowledge provide information about relationships between the variable to be forecasted and explanatory variables. Since causal models can relate planning and decision-making to forecasts, they are useful if one wants to create forecasts that are conditional upon different states of the environment. More important, causal models can be used to forecast the effects of different policies.

Regression analysis involves estimating causal model coefficients from historical data. Models consist of one or more regression equations used to represent the relationship between a dependent variable and explanatory variables. Regression models are useful in situations with few variables and many reliable observations where the causal factors vary independently of one another. Important principles for developing regression (econometric) 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, and (3) discard variables if the estimated relationship conflicts with theory or prior evidence.

Real-world forecasting problems are, however, more likely to involve few observations and many relevant variables. In such situations, the *index method* can be used. Index scores are calculated by adding the values of the explanatory variables, which may be assessed subjectively, for example as zero or one, or may be normalized quantitative data. If there is good prior domain knowledge, explanatory variables may be weighted relative to their importance. Index scores can be used as forecasts of the relative likelihood of an event. They can also be used to predict numerical outcomes, for example by regressing index scores against historical data.

Segmentation is useful when a heterogeneous whole can be divided into homogenous parts that act in different ways in response to changes, and that can be forecasted more accurately than the whole. For example, in the airline industry, price has different effects on business and personal travelers. Appropriate forecasting methods can be used to forecast individual segments. For example, separate regression models can be estimated for each segment. Armstrong (1985, p. 287) reported on three comparative studies on segmentation. Segments were forecasted either by extrapolation or regression analysis. Segmentation improved accuracy for all three studies.

Selection of methods

The Forecasting Method Selection Tree, shown in Figure 2, provides guidance on selecting the best forecasting method for a given problem. The Tree has been derived from evidence-based principles. Guidance is provided in response to the user's answers to questions about the availability of data and state of knowledge about the situation for which forecasts are required. The first question is whether sufficient objective data are available to perform statistical analyses. If not, the forecaster should use judgmental methods.

In deciding among judgmental procedures, one must assess whether the future is likely to be substantially different from the past, whether the situation involves decision makers who have conflicting interests, and whether policy analysis is required. Other considerations affecting the selection process are whether forecasts are made for recurrent and well-known problems, whether domain knowledge is available, and whether information about similar types of problems is available.

—— Figure 2 (Selection Tree) about here ——

If, on the other hand, much objective data are available and it is possible to use quantitative methods, the forecaster first has to assess whether there is useful knowledge about causal relationships, whether cross-sectional or time-series data are available, and whether large changes are involved. In situations with little knowledge about empirical relationships, the next issues are to assess whether policy analysis is involved and whether there is expert domain knowledge about the situation. If there is good prior knowledge about empirical relationships and the future can be expected to substantially differ from the past, the number of variables and presence or absence of inter-correlation between them, and the number of observations determine which causal method to use. For example, regression models that rely on non-experimental data can typically use no more than 3 or 4 variables—even with massive sample sizes. For problems involving many causal variables, variable weights should not be estimated from the dataset. Instead it is useful to draw on independent sources of evidence (such as empirical studies and prior expert knowledge) for assessing the impact of each variable on the situation.

The Forecasting Method Selection Tree provides guidance but on its own, the guidance is not comprehensive. Forecasters may have difficulty identifying the conditions that apply. In such situations, one should use different methods that draw on different information and combine their forecasts according to pre-specified rules. Armstrong (2001c) conducted a meta-analysis of 30 studies and estimated that the combined forecast 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. Studies since that meta-analysis suggest that under ideal conditions (many forecasts available for a number of different valid methods and data sources when forecasting for an uncertain situation), the error reductions from combining are much larger. Simple averages are a good starting point but differential weights may be used if there is strong evidence about the relative accuracy of the method. Combining forecasts is especially useful if the forecaster wants to avoid large errors and if there is uncertainty which method will be most accurate.

The final issue is whether there is important information that has not been incorporated in the forecasting methods. This includes situations in which recent events are not reflected in the data, experts possess good domain knowledge about future events or changes, or key variables could not be included in the model. In the absence of these conditions, one should not adjust the forecast. If important information has been omitted and adjustments are needed, one should use a structured approach. That is, provide written instructions, solicit written adjustments, request adjustments from a group of experts, ask for adjustments to be made prior to seeing the forecast, record reasons for the revisions, and examine prior forecast errors.

Forecasting Canon

The Forecasting Canon provides a summary of evidence-based forecasting knowledge, in this case in the form of nine overarching principles that can help to improve forecast accuracy. The principles are often ignored by organizations, so attention to them offers substantial opportunities for gain.

1. Match the forecasting method to the situation

Conditions for forecasting problems vary. No single best method works for all situations. The Forecasting Method Selection Tree (Figure 2) can help identify appropriate forecasting methods for a given problem. The recommendations in the Selection Tree are based on expert judgment grounded in research studies. Interestingly, generalizations based on empirical evidence sometimes conflict with common beliefs about which forecasting method is best.

2. Use domain knowledge

Managers and analysts typically have useful knowledge about situations. While this domain knowledge can be important for forecasting, it is often ignored. Methods that are not well designed to incorporate domain knowledge include exponential smoothing, stepwise regression, data mining, and neural networks.

Managers' expectations are particularly important when their knowledge about the direction of the trend in a time series conflicts with historical trends in the data (called "contrary series"). If one ignores domain knowledge about contrary series, large errors are likely.

A simple rule can be used to obtain much of the benefit of domain knowledge: when one encounters a contrary series, do not extrapolate a trend. Instead, extrapolate the latest value—this approach is known as the naive or no-change model.

3. Structure the problem

One of the basic strategies in management research is to break a problem into manageable pieces, solve each piece, then put them back together. This decomposition strategy is effective for forecasting, especially when there is more knowledge about the pieces than about the whole. Decomposition is particularly useful when the forecasting task involves extreme (very large or very small) numbers.

When contrary series are involved and the components of the series can be forecasted more accurately than the global series, using causal forces to decompose the problem increases forecasting accuracy. For example, to forecast the number of people who die on the highways each year, forecast the number of passenger miles driven (a series that is expected to grow) and the death rate per million passenger miles (a series that is expected to decrease) and then multiply.

4. Model the experts' forecasts

Expert systems represent forecasts made by experts and can reduce the costs of repetitive forecasts while improving accuracy. However, expert systems are expensive to develop.

An inexpensive alternative to expert systems is judgmental bootstrapping. The general proposition borders on the preposterous; it is that a simple model of the man will be more accurate than the man. The reasoning is that the model applies the man's rules more consistently than the man can.

5. Represent the problem realistically

Start with the situation and develop a realistic representation. This generalization conflicts with common practice, in which one starts with a model and attempt to generalize to the situation. Realistic representations are especially important when forecasts based on unaided judgment fail. Simulated interaction is especially useful for developing a realistic representation of a problem.

6. Use causal models when you have good information

Good information means that the forecaster (1) understands the factors that have an influence on the variable to forecast and (2) possesses enough data to estimate a regression model. To satisfy the first condition, the analyst can obtain knowledge about the situation from domain knowledge and from prior research. Thus, for example, an analyst can draw upon quantitative summaries of research (meta-analyses) on price or advertising elasticities when developing a sales-forecasting model. An important advantage of causal models is that they reveal the effects of alternative decisions on the outcome, such as the effects of different prices on sales. Index models are a good alternative when there are many variables and insufficient data for regression analysis.

7. Use simple quantitative methods

Complex models are often misled by noise in the data, especially in uncertain situations. Thus, using simple methods is important when there is much uncertainty about the situation. Simple models are easier than complex models to understand and less prone to mistakes. They are also more accurate than complex models when forecasting for complex and uncertain situations—which is the typical situation for the social sciences.

8. Be conservative when uncertain

One should make conservative forecasts for uncertain situations. For cross-sectional data, this means staying close to the typical behavior (often called the "base rate"). In time series, one should stay close to the historical average. If the historical trend is subject to variations, discontinuities, and reversals, one should be cautious with extrapolating the historical trend. Only when a historical time series show a long steady trend with little variation should one extrapolate the trend into the future.

9. Combine forecasts

Combining is especially effective when different forecasting methods are available. Ideally, one should use as many as five different methods, and combine their forecasts using a predetermined mechanical rule. Lacking strong evidence that some methods are more accurate than others, one should use a simple average of forecasts.

Conclusion

This entry gives an overview of methods and principles that are known to reduce forecast error. The Forecasting Method Selection Tree provides guidance for which method to use under given conditions. The Forecasting Canon can be used as a simple checklist to improve forecast accuracy. Further information and support for evidence-based forecasting is available from the *Principles of Forecasting* handbook and from the ForecastingPrinciples.com Internet site.

References

Armstrong, J. S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error, *International Journal of Forecasting*, 22, 583-598.

Armstrong (2001a). Judgmental bootstrapping: Inferring experts' rules for forecasting. In: J. S. Armstrong, Editor, *Principles of forecasting*, Kluwer Academic Publishers, Boston, pp. 171-192.

Armstrong (2001b). Extrapolation for time-series and cross-sectional data. In: J. S. Armstrong, Editor, *Principles of forecasting*, Kluwer Academic Publishers, Boston, pp. 217-243.

Armstrong (2001c). Combining forecasts. In: J. S. Armstrong, Editor, *Principles of forecasting*, Kluwer Academic Publishers, Boston, pp. 417-440.

Armstrong, J. S. (1985). *Long-range forecasting*, John Wiley, New York.

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

Collopy, F., Adya, M. & Armstrong, J. S. (2001). Expert systems for forecasting. In: J. S. Armstrong, Editor, *Principles of forecasting*, Kluwer Academic Publishers, Boston, pp. 285-300.

Green, K. C. & Armstrong, J. S. (2007). Structured analogies for forecasting. *International Journal of Forecasting*, 23, 365-376.

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

Green, K. C., Armstrong, J. S. & Graefe, A. (2007). Methods to elicit forecasts from groups: Delphi and prediction markets compared, *Foresight - The International Journal of Applied Forecasting*, 8, 17-20.

MacGregor, D. G. (2001). Decomposition in judgmental forecasting and estimation. In: J. S. Armstrong, Editor, *Principles of forecasting*, Kluwer Academic Publishers, Boston, pp. 107-124.

Makridakis, S., Andersen, S., Carbone, R. Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & R. Winkler (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting*, 1, 111–153.

Makridakis, S. & M. Hibon (2000). The M-3 competition: results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

Rowe, G. & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. In: J. S. Armstrong, Editor, *Principles of forecasting*. Kluwer Academic Publishers: Boston, pp. 125–144.