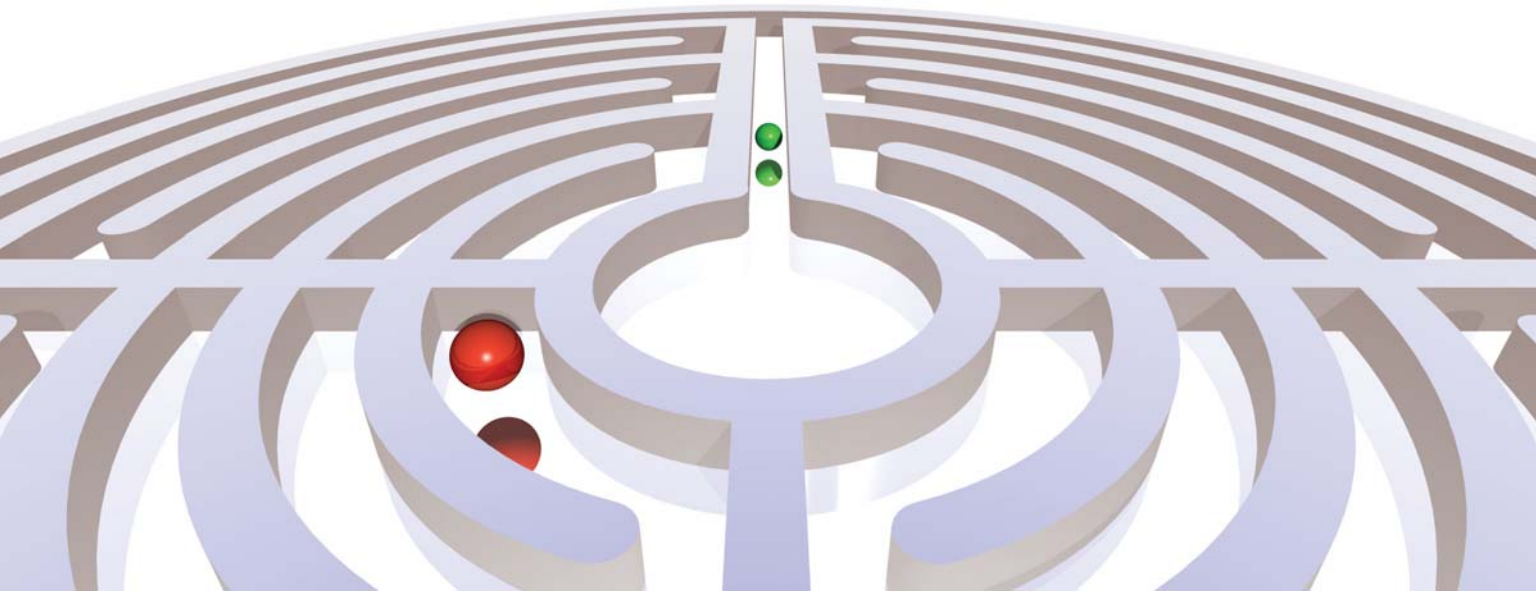


Volume 1 Issue 1 June 2005

FORESIGHT

The International Journal of Applied Forecasting



SPECIAL FEATURE

When and How Should Statistical
Forecasts be Judgmentally Adjusted?

THE FORECASTING CANON

Ways to Improve Forecast Accuracy

A PUBLICATION OF THE INTERNATIONAL INSTITUTE OF FORECASTERS





FORECASTING PRINCIPLES AND METHODS

THE FORECASTING CANON: NINE GENERALIZATIONS TO IMPROVE FORECAST ACCURACY by J. Scott Armstrong

Preview: Using findings from empirically-based comparisons, Scott develops nine generalizations that can improve forecast accuracy. He finds that these are often ignored by organizations, so that attention to them offers substantial opportunities for gain. In this paper, Scott offers recommendations on how to structure a forecasting problem, how to tap managers' knowledge, and how to select appropriate forecasting methods.



J. Scott Armstrong is Professor of Marketing at the Wharton School, University of Pennsylvania. He is a founder of the Journal of Forecasting, the International Journal of Forecasting, and the International Symposium on Forecasting. He is the creator of the Forecasting Principles website, (forecastingprinciples.com) and editor of Principles of Forecasting: A Handbook for Researchers and Practitioners. He is one of the first six Honorary Fellows of the International Institute of Forecasters, and was named Society for Marketing Advances/JAI Press Distinguished Marketing Scholar for 2000.

Using findings from empirical comparisons, I have developed nine generalizations to improve forecast accuracy:

1. Match the forecasting method to the situation
2. Use domain knowledge
3. Structure the problem
4. Model experts' forecasts
5. Represent the problem realistically
6. Use causal models when you have good information
7. Use simple quantitative methods
8. Be conservative when uncertain
9. Combine forecasts

Although these generalizations offer substantial benefits, they are often ignored.

project to develop a new method. The existing forecasting method was primitive: an accountant, aided by a ruler and a calendar, looked at each time series and made a judgmental extrapolation.

At the time, little had been written about forecasting methods. However, drawing upon Brown's (1959) path-breaking book, I developed an exponential smoothing program with seasonal adjustment. Because there was no forecasting software, I wrote a program. The new method provided more accurate forecasts and it was less expensive than the existing method. I was unprepared for the next problem, however; it was not considered proper to change the current process, so the company did not use the program until the accountant retired a few years later.

The world has changed since then. In addition to forecasting software, we now have access to much research on forecasting. Analyses that compare empirical studies are especially important for determining which methods work best in given situations.

Using this evidence-based approach to forecasting, I developed nine generalizations for improving accuracy. These generalizations have been cleverly hidden among thousands of academic papers, so it would not be surprising if some of them have escaped your notice.

I started my career as an industrial engineer at Eastman Kodak in 1960. Noticing that the company had a problem in short-term forecasting for production, I embarked on a

1. Match the forecasting method to the situation.

If you give a forecasting problem to consultants, they will probably use the same method they use for all their forecasting problems. This habit is unfortunate because the conditions for forecasting problems vary. No single best method works for all situations.

To match forecasting methods to situations, I developed a selection tree (See Figure 1). You can describe your problem and use the tree to find which of 17 types of forecasting methods is appropriate. The selection tree is available in hypertext form at forecastingprinciples.com. That allows people to drill down to get details about the methods and to learn about resources such as research, software, and consultants.

Many of the recommendations in the selection tree are based on expert judgment. Most of them are also grounded in research studies. Interestingly, the generalizations based on empirical evidence sometimes conflict with common beliefs about which method is best.

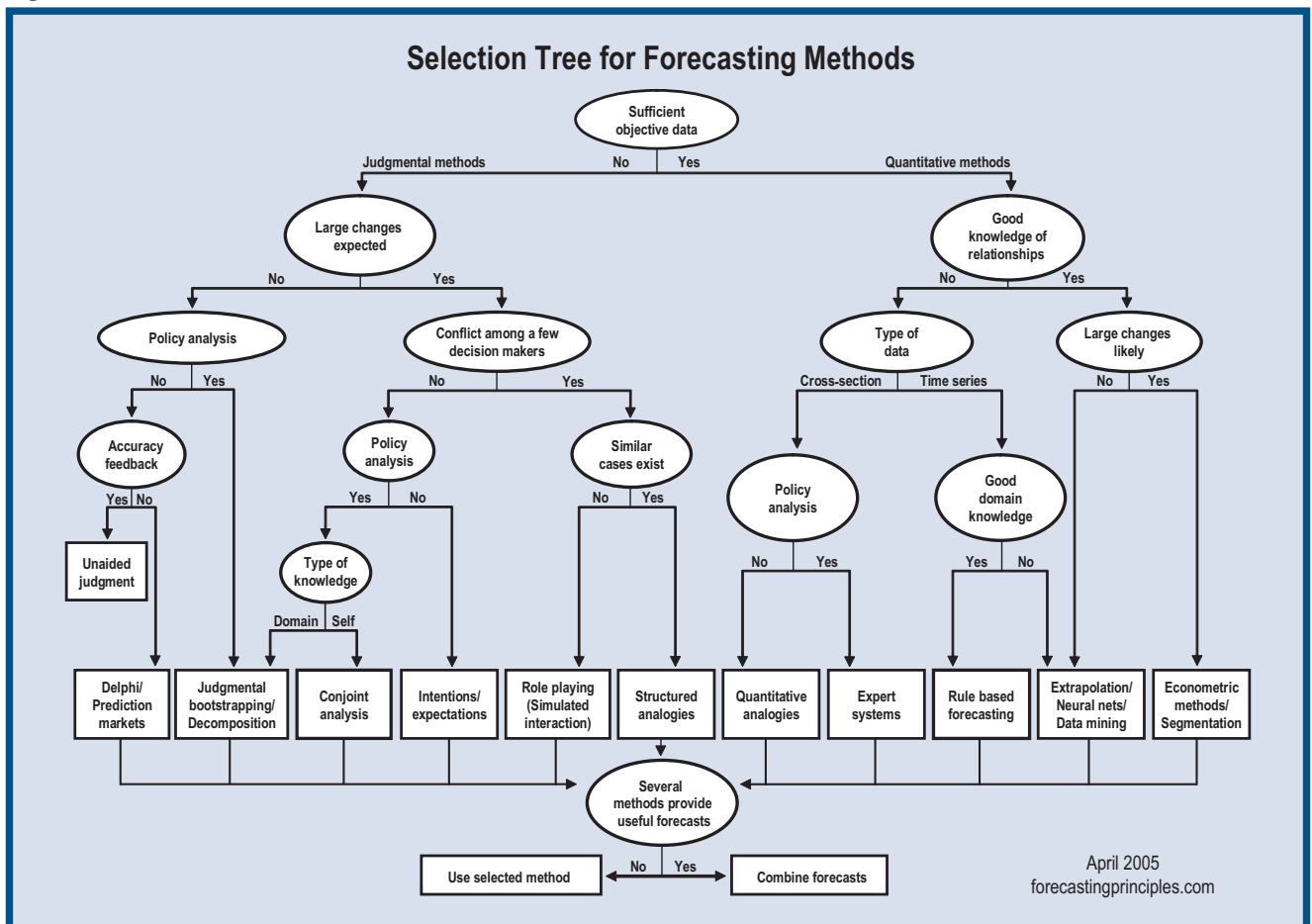
2. Use domain knowledge.

Managers and analysts typically have useful knowledge about situations. For example, they might know a lot about the automobile business. While this domain knowledge can be important for forecasting, it is often ignored. Such methods as exponential smoothing, Box-Jenkins, stepwise regression, data mining, and neural nets seldom incorporate domain knowledge.

Research on the use of domain knowledge has been growing rapidly in recent years. Armstrong and Collopy (1998) found 47 studies on this topic published from 1985 to 1998. These studies provided guidance on how to use judgment most effectively.

One useful and inexpensive way to use managers' knowledge is based on what we call causal forces. Causal forces can be used to summarize managers' expectations about the direction of the trend in a time series. Will the underlying causal forces cause the series to increase or to decrease? Managers' expectations are particularly

Figure 1. Selection Tree



important when their knowledge about causal forces conflicts with historical trends, a situation that we call contrary series. For example, assume that your company has recently come out with a product that will steal substantial sales from one of its existing products whose sales had been increasing. You are shifting your marketing support away from this older product in favor of the new product. The older product represents a contrary series because the historical trend is up, but the expected future trend is down. Forecasts of contrary series by traditional methods usually contain enormous errors.

Causal forces play an important but complicated role in rule-based forecasting, a method for selecting and weighting extrapolation methods (Collopy and Armstrong, 1992). However, you can use a simple rule to obtain much of the benefit of this domain knowledge: when you encounter a contrary series, do not extrapolate a trend. Instead, extrapolate the latest value (the so-called naive or no-change model). When we tested this rule on a large dataset known as the M-competition data (Makridakis et al., 1982) along with data from four other datasets, we reduced errors by 17 percent for one-year-ahead forecasts and over 40 percent for six-year-ahead forecasts (Armstrong and Collopy, 1993).

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 things back together. This strategy is effective for forecasting, especially when you know more about the pieces than about the whole. Thus, to forecast sales, decompose by

- level, trend, and seasonality,
- industry sales and market share for your brand,
- constant dollar sales and inflation, and/or
- different product lines.

These approaches to decomposition can produce substantial improvements in accuracy. For example, in forecasts over an 18-month horizon for 68 monthly economic series from the M-competition, Makridakis et al. (1984, Table 14) showed that seasonal decomposition reduced forecast errors by 23 percent.

MacGregor (2001) showed that decomposition improves the accuracy of judgmental forecasts when the task involves

extreme (very large or very small) numbers. He decomposed 15 problems from three studies, reducing average error by about one-half that of the global estimate.

Forecasting problems can also be structured by causal forces. When contrary series are involved and the components of the series can be forecast more accurately than the global series, decomposing by causal forces improves forecast accuracy (Armstrong, Collopy and Yokum, 2005). 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. When we tested this procedure on five time series that clearly met the conditions, we reduced forecast errors by two-thirds. In addition, for four series that partially met the criteria, we reduced the errors by one-half.

4. Model experts' forecasts.

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

Judgmental bootstrapping offers an inexpensive alternative to expert systems. In this method, you make a statistical inference of a judge's model by running a regression of the judgmental forecasts against the information that the forecaster used. Almost all judgmental bootstrapping models get boiled down to four or fewer variables. 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.

Judgmental bootstrapping provides greater accuracy than judges' forecasts (Armstrong, 2001a). It was superior to unaided judgment (the normal method for these situations), in eight of the 11 comparisons, with two tests showing no difference, and one showing a small loss. All of these comparisons used cross-sectional data.

Judgmental bootstrapping has additional advantages because it shows experts how they are weighting various factors. This knowledge can help them to improve their judgmental forecasting. For example, with respect to personnel selection, bootstrapping might reveal that

managers consider factors (such as height) that are not relevant to the job. Bootstrapping also allows forecasters to estimate the effects of changing key variables when they have no historical data on such changes and thus avoid assessing them with econometric methods.

Although fairly inexpensive, judgmental bootstrapping is seldom used by practitioners. Perhaps it is because the results violate our common sense, or perhaps it is because we do not like to think that a computer can make better forecasts than we can.

5. Represent the problem realistically.

Start with the situation and develop a realistic representation. This generalization conflicts with common practice, in which we start with a model and attempt to generalize to the situation. This practice helps to explain why game theory, a mathematical model used to model and predict the behavior of adversaries in a conflict, has had no demonstrable value for forecasting (Green, 2005).

Realistic representations are especially important when forecasts based on unaided judgment fail, as they do when forecasting decisions are made in conflict situations. Simulated interaction, a type of role-playing in which two or more parties act out interactions, is a realistic way to portray situations. For example, to predict how a union will react to a company's potential offer in a negotiation, people play the two sides as they decide whether to accept this offer. Compared to expert judgment, simulated interactions reduced forecast errors 44 percent in the eight situations Green (2005) studied.

Another approach to realism is to identify analogous situations. Green and Armstrong (2004), using eight conflict situations, found that a highly structured approach to using analogies reduced errors by 20 percent. When the experts could think of two or more analogies, the errors dropped by more than 40 percent.

6. Use causal models when you have good information.

By good information, I mean enough information to understand the factors that affect the variable to be forecast, and enough data to develop a causal (econometric) 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 pricing or advertising elasticities when developing a sales-forecasting model. Such information is not used for data mining which might account for the fact that, to date, there have been no comparative studies showing that data mining improves forecast accuracy.

Allen and Fildes (2001) present evidence showing that quantitative econometric models are more accurate than noncausal methods, such as exponential-smoothing models. Quantitative econometric models are especially important for forecasting situations involving large changes.

Fair (2002) provides a good overview of econometric methods and illustrates them with a series of practical problems. However, he ignores an important reason for using econometric models, which is to examine the effects of policy variables. Causal models allow one to see the effects of alternative decisions, such as the effects of different prices on sales.

7. Use simple quantitative methods.

One of the primary conclusions drawn from the series of M-competition studies, which involved thousands of time series, was that beyond a modest level, complexity in time-series extrapolation methods produced no gains (Makridakis and Hibon, 2000). Based on evidence summarized by Armstrong (1985, pp. 225-235), this conclusion also applies to econometric studies. Furthermore, although researchers have made enormous efforts to develop models of how ownership of new consumer goods spreads through a population, Meade and Islam (2001) concluded that simple diffusion models are more accurate than complex ones.

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 to understand, less prone to mistakes, and more accurate than complex models.

8. Be conservative when uncertain.

The many sources of uncertainty make forecasting difficult. When you encounter uncertainty, make conservative forecasts. In time series, this means staying close to an

historical average. For cross-sectional data, stay close to the typical behavior (often called the base rate).

When a historical time series shows a long steady trend with little variation, you should extrapolate the trend into the future. However, if the historical trend is subject to variations, discontinuities, and reversals, you should be less willing to extrapolate the historical trend. Gardner and McKenzie (1985) developed and tested a method for damping trends in extrapolation models. In a study based on 3003 time series from the M-competition, damped trends with exponential smoothing reduced forecast errors by seven percent when compared with traditional exponential smoothing (Makridakis and Hibon, 2000). The U.S. Navy implemented a program for 50,000 items, reducing inventory investment by seven percent, a \$30 million savings (Gardner, 1990). Some software packages now allow estimation of damped trends for exponential smoothing.

Miller and Williams (2004), using time series from the M-competition series, developed a procedure for damping seasonal factors. When there was more uncertainty in the historical data, they used smaller seasonal factors (e.g., multiplicative factors were drawn towards 1.0). Their procedures reduced forecast errors by about four percent. Miller and Williams provide freeware at forecastingprinciples.com.

9. Combine forecasts.

Researchers have recommended combining forecasts for over half a century. In surveys of forecasting methods, many organizations claim to use combined forecasts. I suspect, however, that most organizations use them in an informal manner and thus miss most of the benefit.

You can typically improve accuracy by using a number of experts. Research support for this recommendation goes back to studies done in the early 1900s. A group of experts usually possesses more knowledge than an individual expert. Unfortunately, however, much of that benefit is forfeited when experts make forecasts in traditional meetings. Simple averages of independent judgmental forecasts, however, can lead to improved forecasts. In a recent study of forecasting decisions in eight conflict situations, Green (2005) found that a combination of judgmental forecasts from simulated interactions reduced

error by 67 percent compared to forecasts from single interaction trials.

In addition to simple averaging, two related techniques, Delphi and prediction markets, can improve forecasts. They reduce biases because practitioners make the forecasts objectively following a preset mechanical rule (e.g., take the median) to combine anonymous forecasts. With both of these techniques, and especially with prediction markets, forecasters are motivated to produce accurate forecasts.

In the Delphi procedure, an administrator obtains at least two rounds of independent forecasts from experts. After each round, the experts are informed about the group's prediction and, in some cases, about reasons. In their review of research on the Delphi procedure, Rowe and Wright (2001) found that Delphi improved accuracy over traditional groups in five studies, worsened it in one study, and tied with traditional methods in two studies. Few of the researchers estimated the error reductions, although one found an error reduction of about 40 percent. As might be expected, when the panelists made forecasts in areas in which they had no expertise, Delphi was of no value.

Prediction markets allow anyone in a given set of people to bet on the outcome of a situation. Wolfers and Zitzewitz (2004) describe and summarize the evidence on prediction markets (also known as betting markets and information markets). The evidence suggests that prediction markets are more accurate than voter intention polls. In addition, some unpublished studies suggest that companies can use them to produce accurate sales forecasts. However, to date, researchers have conducted no large-scale empirical studies to compare forecasts from prediction markets with those from traditional groups.

These methods for using judgments assume that the individuals can make useful forecasts. Typically they can, but not always. Green and Armstrong (2005) in research on eight conflict situations, found that unaided predictions by experts were little different from predictions based on chance. These methods assume that at least some of the individuals have conducted relevant analyses. For example, I show people that the thickness of a piece of paper increases as I fold it in half about six times; I then ask them to forecast, *without doing any calculations*, the thickness of the paper when folded in half 40 times. (Presumably, everyone would know how to do the calculations.)

Combining forecasts is of little help for this problem, because all forecasts are too small; the typical answer misses by a factor of about one million. You can easily replicate this demonstration.

Combining can also be used for other methods. In a quantitative review of 30 studies, combining forecasts improved accuracy in each study compared with the typical method (Armstrong, 2001b). The gains ranged from three to 24 percent with an average error reduction of 12 percent. In some cases, the combined forecast was better than any of the individual methods.

Combining is especially effective when different forecasting methods are available. Ideally, 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, an equally weighted average of the forecasts should work well.

To demonstrate the value of combining forecasts from different methods, Cuzán, Armstrong and Jones (2005) applied it to the 2004 U.S. presidential election. This situation was ideal for combining because there were a number of methods to combine and a large number of variables that might influence voter choice. The combined forecast, the Pollyvote, was based on an equally weighted average of polls (which were themselves combined), econometric methods (also combined), Delphi, and a prediction market. The Pollyvote reduced the forecast error by about half compared to the typical forecast from polls or from a prediction market. The Pollyvote was also more accurate than the best of the component methods. (See RECOGNITION FOR FORECASTING ACCURACY on pages 51-52 in this issue of FORESIGHT, where the authors discuss forecasting the 2004 U.S. presidential election.)

Conclusion

Comparative empirical studies have led to generalizations that can produce forecasts that are more accurate than those from commonly used methods. Few organizations capitalize on this knowledge. Why do they fail to act? Is the problem the same one I faced long ago at Eastman Kodak? Must we wait for the current analysts to retire? Or is there another way?

References

(Papers by Armstrong are available in full text at forecastingprinciples.com)

Allen, Geoff & Fildes, Robert (2001). Econometric forecasting. In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (2001a). Judgmental bootstrapping. In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (2001b). Combining forecasts. In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (1985). *Long-Range Forecasting*. New York: John Wiley.

Armstrong, J. S. & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. In: G. Wright & P. Goodwin (Eds.), *Forecasting with Judgment*. New York: John Wiley & Sons Ltd., 269-293. *International Journal of Forecasting*, 15, 345-346.

Armstrong, J. S. & Collopy, F. (1993). Causal forces: Structuring knowledge for time-series extrapolation, *Journal of Forecasting*, 12, 103-115.

Armstrong, J. S., Collopy, F. & Yokum, J.T. (2005). Decomposition by causal forces: A procedure for forecasting complex time series, *International Journal of Forecasting*, 21, 25-36.

Brown, Robert G. (1959). *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.

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

Cuzán, Alfred G., Armstrong, J.S. & Jones, Randall J., Jr. (2005). Combining methods to forecast the 2004 presidential election: The pollyvote, Working paper January.

Fair, Ray C. (2002). *Predicting Presidential Elections and Other Things*. Stanford, CA: Stanford University Press.

Gardner, E.S., Jr. (1990). Evaluating forecast performance in an inventory control system, *Management Science*, 36, 490-499.

Gardner, Everette S. & McKenzie, E. (1985). Forecasting trends in time series, *Management Science*, 31, 1237-1246.

Green, Kesten C. (2005). Game theory, simulated interaction, and unaided judgment for forecasting decisions in conflicts: Further evidence, *International Journal of Forecasting*, 21, forthcoming. In full text at conflictforecasting.com.

Green, Kesten C. & Armstrong, J.S. (2005). Value of expertise for forecasting decisions in conflicts, Working paper, Department of Marketing, The Wharton School, University of Pennsylvania.

Green, Kesten C. & Armstrong, J.S. (2004). Structured analogies for forecasting, Working paper, Department of Marketing, The Wharton School, University of Pennsylvania.

MacGregor, Donald G. (2001). Decomposition for judgmental forecasting and estimation. In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

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

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1984). *The Forecasting Accuracy of Major Time-series Methods*. Chichester: John Wiley.

Makridakis, Spyros & Hibon, M. (2000). The M3 competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

Meade, Nigel & Islam, T. (2001). Forecasting the diffusion of innovations: Implications for time series extrapolation, In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Miller, Don M. & Williams, Dan (2004). Shrinkage estimators for damping X12-ARIMA seasonals, *International Journal of Forecasting*, 20, 529-549.

Rowe, Gene & Wright, George (2001). Expert opinions in forecasting: The role of the Delphi technique, In: J. S. Armstrong (Ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Wolfers, Justin & Zitzewitz, Eric (2004). Prediction markets, *Journal of Economic Perspectives*, 18 (2) 107–126.

Contact Info:

J. Scott Armstrong
The Wharton School
University of Pennsylvania
armstrong@wharton.upenn.edu

FORESIGHT subscription info at



[more on page 4](#)

FORESIGHT online