Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Forecasting 22 (2006) 583–598
international journal of forecasting
www.elsevier.com/locate/ijforecast

Findings from evidence-based forecasting: Methods for reducing forecast error

J. Scott Armstrong

Wharton School, University of Pennsylvania, United States

Abstract

Empirical comparisons of reasonable approaches provide evidence on the best forecasting procedures to use under given conditions. Based on this evidence, I summarize the progress made over the past quarter century with respect to methods for reducing forecasting error. Seven well-established methods have been shown to improve accuracy: combining forecasts and Delphi help for all types of data; causal modeling, judgmental bootstrapping and structured judgment help with cross-sectional data; and causal models and trend-damping help with time series data. Promising methods for cross-sectional data include damped causality, simulated interaction, structured analogies, and judgmental decomposition; for time series data, they include segmentation, rule-based forecasting, damped seasonality, decomposition by causal forces, damped trend with analogous data, and damped seasonality. The testing of multiple hypotheses has also revealed methods where gains are limited: these include data mining, neural nets, and Box–Jenkins methods. Multiple hypotheses tests should be conducted on widely used but relatively untested methods such as prediction markets, conjoint analysis, diffusion models, and game theory.

Keywords: Box–Jenkins; Causal forces; Causal models; Combining forecasts; Complex series; Conjoint analysis; Contrary series; Damped seasonality; Damped trend; Data mining; Delphi; Diffusion; Game theory; Judgmental decomposition; Multiple hypotheses; Neural nets; Prediction markets; Rule-based forecasting; Segmentation; Simulated interaction; Structured analogies

This paper summarizes what has been learned over the past quarter century about the accuracy of forecasting methods. It relies on empirical studies that compare 'multiple hypotheses' (two or more reasonable hypotheses). This method of reasonable alternatives implies that the current method is included along with other leading methods. Ideally, the hypotheses should specify the conditions under which the findings apply. I refer to this approach as *multiple hypotheses* and to the findings as *evidence-based*.

1. Evidence-based findings

In judging progress in a field, one might look at new methods and develop a rationale on why they should be useful. Consider an analogy to medical research: one could develop new treatments in the lab based on reasoning about what treatments should be

E-mail address: armstrong@wharton.upenn.edu.

0169-2070/$ - see front matter © 2006 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2006.04.006

most effective and then have them judged by experts. In a like manner, Fildes (2006-this issue) examined the most influential new treatments in forecasting. Peer review has supported these approaches. Is this sufficient?

Continuing with the analogy to medicine, Avorn (2004, p. 23) reports the following, which I have paraphrased: "In a former British colony, most healers believed the conventional wisdom that a distillation of fluids extracted from the urine of horse, if dried to a powder and fed to aging women, could ... preserve youth and ward off a variety of diseases." The preparation became very popular. Many years later, experimental studies concluded that the treatment had little value and that it caused tumors and blood clots. The former colony is the United States and the drugs were hormone replacement products. The treatment seemed to work because those who used the drug tended to be healthier than those who did not. This was because it was used by people who were more interested in taking care of their health.

I have little faith in the value of forecasting methods until they have been empirically tested. Popular techniques have often failed when subjected to testing. So in this paper, I examine methods that have been empirically tested against other methods. As is the case for most research in the social and managements sciences, only a small percentage of papers are concerned with evaluation.

I looked primarily for studies that used real data to compare the ex ante forecasting accuracy of alternative methods. Where possible, I relied upon published reviews and meta-analyses.

My search for evidence-based findings was intended to include all types of forecasting methods. Using the forecasting methodology tree at forecastingprinciples.com, I examined 17 basic methods: role playing, intentions/expectations surveys, conjoint analysis, prediction markets, Delphi, structured analogies, game theory, decomposition, judgmental bootstrapping, expert systems, extrapolation models, data mining, quantitative analogies, neural nets, rule-based forecasting, causal models, and segmentation. Brief summaries of these methods are available at forecastingprinciples.com with additional details in Armstrong (2001a).

While this review focuses on the first 25 years of the International Institute of Forecasters (from its founding in 1981), many of the advances are built upon earlier work. Earlier contributions, such as the classical decomposition of time series (mean, trend, and seasonality) are not discussed if I was unable to obtain evidence from the past 25 years that related to the use of the methods.

The initial base of findings is drawn from Armstrong (2001a). In that book, 39 academic researchers in forecasting summarized evidence-based principles in their areas. They were supported by 123 reviewers in an effort to ensure that all relevant evidence on the principles had been included. The principles were initially posted on an open website, forecastingprinciples.com, and appeals were made for peer review as to any information that had been overlooked.

I began to update the review in early 2005 by searching the literature, contacting key researchers, and requesting help through various email lists (e.g., the Associate Editors of the *International Journal of Forecasting*, and the authors and reviewers of the *Principles of Forecasting* book). An early version of this paper was presented as a keynote address at the International Symposium on Forecasting in 2005 along with an appeal for peer review. Drafts were posted for months on forecastingprinciples.com along with a request for reviews. I also asked a number of experts to act as reviewers of this paper. I am indebted to the 23 reviewers who provided substantive contributions to the paper as well as to others who made smaller contributions. Some of these reviewers read more than one version of the paper.

Advances in methods for improving forecast accuracy in the past 25 years are summarized below. The review begins with methods that are well established, moves to "promising methods," proceeds to those that have been tested but found to offer only limited gains, and concludes with methods that have been widely used but not well-tested.

Within each of these areas, the methods are organized by those that apply to all types of data, followed by those relevant primarily for cross-sectional data, and then those applicable to time-series data. In assessing improvements, I sought evidence on the percentage reduction in the absolute ex ante forecast error. When there was little evidence of error reduction, I reported on the percent of the time the specified method improved accuracy. In addition to examining evidence on the accuracy of the