A brief history of forecasting competitions

Abstract

Forecasting competitions are now so widespread that it is often forgotten how controversial they were when first held, and how influential they have been over the years. I briefly review the history of forecasting competitions, and discuss what we have learned about their design and implementation, and what they can tell us about forecasting. I also provide a few suggestions for potential future competitions, and for research about forecasting based on competitions.

Keywords: evaluation, forecasting accuracy, Kaggle, M competitions, neural networks, prediction intervals, probability scoring, time series

Prediction competitions go back millennia; for example, rival diviners in ancient Greece competed to predict the future more accurately (Raphals 2013, p124). However, for general time series forecasting (i.e., predicting the future of regularly observed data over time), the history is much more limited and goes back only about 50 years. In fact, it wasn't until computers were widely available that it became feasible for forecasting competitions to be held at all.

Time series forecasting competitions have been a feature of the *International Journal of Forecasting* and the *Journal of Forecasting* since the journals were founded in the early 1980s. This strong emphasis on large scale empirical evaluations of forecasting methods, and the need to compare newly proposed methods against existing state-of-the-art methods, has played a large part in pushing researchers to develop new methods that can be shown to work in practice (Fildes & Ord 2002).

Researchers new to forecasting are often surprised to learn how controversial such competitions were when they were first conducted about 50 years ago. I review this controversy in section 1. The influential series of Makridakis competitions are discussed in section 2, and other forecasting competitions are described in section 3. Finally, I provide a few comments on the future of forecasting competitions, and research about forecasting competitions, in section 4. I do not cover forecasting competitions that are not based around time series data.

1 Early controversy

The earliest forecasting competitions were between methods rather than people. It was not feasible, given the communication tools available at the time, to conduct a large-scale forecasting competition involving many entrants spread around the world. So the first few competitions were by individual researchers comparing the accuracy of several methods applied to multiple time series. I only include the first two of these. From 1980 onwards, my scope is restricted to competitions involving multiple entrants.

Nottingham studies

The earliest non-trivial study of time series forecast accuracy was probably by David Reid as part of his PhD at the University of Nottingham (Reid 1969). Building on his work, Paul Newbold and Clive Granger conducted a study of forecast accuracy involving 106 time series (Newbold &