

## Probability Forecasting

Effort on the evaluation of *probability forecasts* continues. This work has been conducted in the past, jointly with my former student John Kling, under the heading of “prequential analysis”, which argues that a major purpose of model-building is to provide “good” out-of-sample forecasts, not necessarily to provide good fit with prior theory (see Kling and Bessler *Jo Business* 1989 and *Jo Royal Stat Soc series c* 1990). For the most part, previous work on prequential analysis has centered on probability calibration as the measure of forecasting “goodness”. Other metrics that consider “sorting” as well as calibration, known as proper scoring rules, have been discussed in the literature. These rules have been proposed for both motivation and evaluation of probability forecasts. For what appears to be the first empirical test of scoring rules as motivational devices see Nelson and Bessler *AJAE* 1989. Our recent focus has been on the Brier Score (a proper scoring rule) and its covariance partition to assess performance (evaluation). A former student, Robert Ruffley, and I investigate these ideas using U.S. stock market data (Bessler and Ruffley *Applied Economics* 2004). We show that “goodness” as measured by tests of calibration does not necessarily imply particularly useful forecasts. The Brier Score, with its partition, offers a more discriminating metric of performance. Gabriel Casillas used this partition to demonstrate how probability forecasts by the Bank of England can be improved to enhance decision making related to price inflation and aggregate output (Casillas and Bessler *Journal of Policy Modeling* 2006). A key contribution in Casillas’ work is the discussion of the use of a “shadow” forecaster, an alternative forecaster to whom comparisons can be made to determine “goodness”. This offers a possible solution to issues arising in the motivation of probability statements made by public sector assessors (issues that proper scoring rules seem inadequate to address). More recent focus has been on the use of the Brier score to evaluate the probability forecasts of discrete choice models (logit, probit, etc.). Here my co-authors and I offer evidence that such “scoring” offers an improvement relative to the usual “threshold percent correct” measure (the latter is given criticism in Stock, J.H. and M.W. Watson, 2007, *Introductory Econometrics*, Addison-Wesley, Boston).

Our study of conditional probability forecasts following the ideas outlined and briefly applied in Bessler and Kling (*AJAE* 1989) is contemplated as future work in this area.