A Suggested Modification of the Brier and Allen Score for Probability Forecasts

MERLE P. WOODALL
Belknap College, Center Harbor, N. H.
7 November 1966

Is it too late to comment on Prof. Sanders' article, On Subjective Probability Forecasting (Sanders, 1963)? Contrary to Sanders' conclusion with respect to the possibility of a forecast strategy harmful to the utility of the forecast, experience at Belknap College in 1965-66 has shown that there is indeed a method to "play" the Brier and Allen score.

Consider an event whose forecast probability is very low (or high). At first glance it might appear rational to forecast 0/10 if the estimated probability is less than 0.5/10, and 1/10 if greater than 0.5/10.

But since the departure of forecast from observed is squared in the Brier and Allen score, i.e.,

\[ F = \frac{1}{N} \sum_{i=1}^{N} (f_i - O_i)^2, \]

the forecaster's decision point for the forecast probability statements is not at 0.5 when \( f_i \) is less than completely reliable. For example, if a forecast of 0.9 is incorrect, the score is 0.81, and if a forecast of 1.0 is incorrect, the score is 1.00, a difference of 0.19; on the other hand, if the forecasts are correct the difference in scores is 0.01. In other words, the forecaster is risking a possible loss of 19 points in the forecast score for a possible gain of 1 point, not a good risk when \( f_i \) has some error, the normal situation.

I suspect that most forecasters making probability forecast have long since realized this situation and that this explains in part the forecasts of 1/10 chance of rain, the so-called chicken zero, when there seems to be only the very slightest possibility of rain.

The forecast score can thus be "played" by staying away from the extremes except in very rare cases, and by frequently shading the forecast toward less sharp values, or by considering the forecaster's gains and losses in the score. I submit that this is not always in the forecaster user's interest.

If the quantity \( (f_i - O_i) \) were not squared, I believe the score would be unbiased when \( f_i \) is other than the climatic frequency. The only reason for squaring the difference seems to be statistical habit, which in this case serves no purpose other than giving forecasters an opportunity to beat the system.

REFERENCES