

## NOTES AND CORRESPONDENCE

## Comments on "Nonlinear Utility and the Probability Score"

LAWRENCE A. HUGHES

*National Weather Service, Central Region, NOAA, Kansas City, Mo.*

5 June 1970 and 26 October 1970

I don't see the point of the paper by Winkler and Murphy (1970) in which they discuss scoring systems to maximize the score for different kinds of utility functions (systems) of forecasters. In the design of a forecast verification scheme, one should create a score that relates directly to the *user's* needs, and is a "proper" scoring system, as the *user* is the one being served, not the forecaster. The Brier score serves both these purposes quite well [see Winkler and Murphy (1968) and Murphy (1969)]. Any forecaster utility function (system) that does not tend to improve the Brier score should be and is strongly discouraged.

In the early stages of the Weather Bureau's entry into the precipitation probability forecasting program (during the trial, non-public, stage), and perhaps to a degree today, some forecasters had utility functions (systems) that caused them to have poorer scores than they otherwise would have had. This was mainly because the knowledge of the effect of such systems on the scores was not yet known. However, when verifications are available, especially individual forecaster verifications, and they are accumulated for a reasonably representative period, say six months, forecasters have been observed to gradually change their ways so they could compete in score with their fellow workers. It would take a very strong-minded forecaster to continue to receive poor verification scores, compared to his co-workers and against the advice of his supervisor, in order to adhere to a utility function he felt was more proper. If he continues to hold such a utility function, the efforts should be to get him to change his function and not to revise the scoring system so as to maximize the score under his utility function.

We are thus concerned only with the utility function of the decision maker and not with that of the forecaster, except to help bring the forecaster's function more in line with that of the decision maker if they are not the same. There is no interest or value in changing a scoring system to match the utility function of the forecaster.

I have heard forecasters say that they are trying to put out good forecasts, not play a numbers game and get a better verification score. To this we need to be able to say that playing the numbers game and getting a better score is the same as putting out better forecasts, as a higher score means forecasts that have more utility—more economic value—to users of the forecasts. Because of the excellent works of Murphy and Winkler, and others, we know that we are correct in this statement when the Brier score is the verification score. Our prime need now is to be able to normalize this score so effects of climatology, i.e., precipitation variables of frequency, mean areal coverage, and the trace-measurable amount ratio can be properly allowed for and thus permit a proper comparison of forecasts among stations and among forecasters.

## REFERENCES

- Murphy, A. H., 1969: Measures of the utility of probabilistic predictions in cost-loss ratio decision situations in which knowledge of the cost-loss ratios is incomplete. *J. Appl. Meteor.*, **8**, 863-873.
- Winkler, R. L., and A. H. Murphy, 1968: "Good" probability assessors. *J. Appl. Meteor.*, **7**, 751-758.
- , and —, 1970: Nonlinear utility and the probability score. *J. Appl. Meteor.*, **9**, 143-148.