Viewpoints

CORRESPONDENCE ON "SCORING RULES FOR SUBJECTIVE ASSESSMENTS"

PROFESSOR J. N. MORSE, UNIVERSITY OF DELAWARE, NEWARK, DELAWARE 18711 WRITES:

"KIDD[1] chooses an unfortunate example for assessing scorers. Any recent finance journal will include some reference to the random walk theory of speculative markets, e.g. Sidney S. Alexander[2] "... where trends seem to be observable, they are merely interpretations, read in after the fact, of a process that really follows a random walk."

Every period the stock market will move up or down; half of the assessors will always be correct. By chance a small percentage of them will be able to appear to predict most of the market's movement over several periods. Malkiel[3] reviews the evidence on the independence of stock price relatives. For example, it is known that there is no statistically significant correlation in the performance ranks of mutual funds from year to year.

We must not confuse a *posteriori* curve-fitting with true predictive ability. I am sure that Kidd's scoring rules are of considerable importance in domains where the process being assessed is not a random walk."

J. B. KIDD, University of Aston in Birmingham, replies:

"The use of scoring rules is meant to adjudicate upon assessors who can process perceived data and verbalize a prediction before the event. The paper shows that some people can be seen to predict an event whilst others can not. In no way was the paper a report upon the process of the generation of stock market values. It was upon the process of deciding who may be employable to forecast the magnitude of a stochastic variable. Furthermore, there was an inference drawn that task complexity causes assessors to perform with a negative correlation with respect to increasing complexity."

Professor Morse responds:

"Kidd responds "The paper shows that some people can be seen to predict an event whilst others cannot." My point is that this statement is false. In no way does the example indicate any predictive ability on the part of the assessors.

The philosophical issue is spurious correlation. For example, out of any finite set of assessors, exactly half will be found to predict the gross directionality of subsequent market movements. But, as is evident from references (2) and (3), this result is not replicable. At each repetition, I suspect, a new group of assessors would surface in the prescient half.

I am objecting to the design of the experiment. It lacks internal validity, external validity and replicability.

I admire the vigour of this experimental effort. There are domains of stochastic events that lend themselves to scoring assessors. Markovian processes, where there is dependence in the time series, would be a place where statistically significant scoring of assessors would be fine.

In addition, I agree completely that increasing task complexity decreases the ability to assess the magnitude of a random variable. This is due to the inadequacy of the human brain as an information processor.

Summarizing, I find Kidd's concept interesting and well thought out. I merely object to the choice of the stock market as a relevant area of application."

REFERENCES

[1] JOHN B. KIDD (1975) "Scoring Rules for Subjective Assessments", *Opl Research Quarterly*, vol. **26**, 183–195.

[2] PAUL COOTNER (editor), *The Random Character of Stock Market Prices*, MIT Press, Cambridge, MA, 1967.

[3] BURTON G. MALKIEL, *A Random Walk Down Wall Street*, W. W. Norton and Company, NY, 1973.