Behavioural and Experimental Economics: Discussion Questions

Judicial Error by Groups and Individuals

Authors: Frans van Dijk, Joep Sonnemans, Ed Bauw

Thomas A. Stephens

29 May 2013

1. How would relaxing the normative model's assumption of risk-neutral expected-utility maximisation affect the results?

According to van Dijk, Sonnemans, and Bauw (2012), subjects' estimates of probabilities are relatively accurate, but decisions to convict diverge considerably from optimality for the normative model of risk-neutral expected-utility maximisers. An interesting question, then, is how relaxing this assumption would affect their results.

Expected utility assumes risk aversion, based on the concavity of the utility function in wealth. However, as Rabin and Thaler (2001) point out, with small stakes (as in this experiment), this approximates risk neutrality. In contrast, non-expected utility models (see Starmer 2000 for a survey) often propose risk neutrality even with small stakes. A prominent example is prospect theory (Kahneman and Tversky 1979), which proposes a value function with asymmetric treatment of gains and losses, as opposed to a utility function based on states of wealth, together with probability weighting.

Under van Dijk et al.'s (2012) normative model, the optimal strategy is to convict if the posterior probability that the subject is guilty is 80 per cent or higher. Intuitively, however, risk and loss aversion should increase this threshold, because conviction involves both a higher variance and a larger potential loss. Using the weighting function and parameter estimates suggested by Tversky and Kahneman (1992), conviction is

optimal only if the posterior probability of guilt is at least 94 per cent.[1] Given that van Dijk et al. (2012) find excessively high conviction rates, relative to the normative model, risk and loss aversion are unlikely to be the cause of deviations from optimality.

The excessively high conviction rates in the experiment are inconsistent with rational payoff maximisation, under both expected utility and prospect theory. One speculative explanation might be the framing as judicial decisions. It is plausible that many of the subjects know people who have been victims of crime, or have even been victims themselves. Conversely, being falsely accused of a crime is a relatively rare event. Decisions may therefore have been influenced by beliefs and heuristics from outside.

2. The statistical analysis may have some problems. Given the large number of observations, would it be feasible to do a parametric analysis?

The statistical analysis involves nonparametric comparisons of individual and group averages across cases, either all 30 cases or subsets ranging in size from 8 to 13. It is not a very intuitive approach, and there may be some problems with it. To start with, the observations are averages across cases, but individual decisions in the group treatment would not be independent. However, the comparison of individual (IndGroup) and group (GroupGroup) decisions in the group treatment uses the Wilcoxon signed rank test, which assumes the paired differences are independent. How can this be the case?

Within groups, there are two sources of dependence. The first is the fact that a group decision is the same for all group members, so there is only one independent observation per group. The second is that, over multiple cases, feedback from other group members would affect individual decisions. This is even mentioned in the discussion of learning effects. The implications would depend in part on the ordering of the cases, i.e. fixed or random, but it is not clear from the paper which is the case. Either way, individual averages cannot be viewed as independent. A partial solution might be to use group averages in all cases, even for individual decisions.

[1] The weighting functions for gains and losses, respectively, are $\frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{1/\gamma}}$ and $\frac{p^\delta}{(p^\delta+(1-p)^\delta)^{1/\delta}}$, where p is the probability of the gain or loss and parameter estimates are $\gamma = 0.61$ and $\delta = 0.69$.

Apart from violations of independence, there are problems with the distributional assumptions and hypotheses of the tests used. The Mann-Whitney rank sum test, for example, is a comparison of distributions, with the null being identical distributions and the alternative being that the distributions differ only in location. In comparing distributions of errors in individual and group decisions, it is not clear that it is reasonable to assume identical dispersion. If this assumption does not hold, then rejection of the null implies only that the distributions differ in some way, and not that the locations necessarily differ.

Given the large number of subjects, $99 + 122 = 221$ (in approximately $33 + 40 = 73$ groups), with 30 decisions per subject, an alternative approach might be a parametric regression. This approach would make it possible to handle violations of independence, e.g. by clustering and using fixed or random effects for groups and periods. The proposed learning effects could also be explored in more detail, by comparing treatment effects in earlier and later periods.

References

Kahneman, D. and Tversky, A. (1979). Prospect Theory - Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.

Rabin, M. and Thaler, R. H. (2001). Anomalies - Risk Aversion. *Journal of Economic Perspectives, 15*(1), 219-232.

Starmer, C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature, 38*(2), 332-382.

Tversky, A. and Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

van Dijk, F., Sonnemans, J., and Bauw, E. (2012). Judicial Error by Groups and Individuals. *Tinbergen Institute Discussion Paper, TI 2012-029/1.*