Referee Report on MS 5073

for the Journal of the European Economic Association

Expectation Formation in an Evolving Game of Uncertainty: Theory and New Experimental Evidence

This very interesting paper reports on the results from a rather complicated experiment designed to understand the relationship between subjective and objective probabilities and the factors determining this relationship.

I find the project interesting and ambitious, though I wonder how the subjects found it. It looks difficult for them to understand (I do not have the Instructions) and wonder how long it took, though I do note that (some) subjects made a lot of money!

The authors, when they say *"We avoid the danger that risk aversion will lead individuals to misrepresent their probability perceptions … by separating their choice behavior from their stated probabilities, and only incentivizing the former"* make the important point that the elicited subjective probabilities that they use are un-incentivised, whereas their decisions (how much to wager and how much to invest, and whether to pull out of the races) are incentivised. I take the point, but note that there are ways to get round the problem that they state: I have recently been reading a paper *A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs* by Karl Schlag, James Tremewan and Joel van der Weele, which discusses a variety of different elicitation methods, some of them taking into account non-risk-neutral behaviour. But irrespective of this point – and let me assume for the moment that their un-incentivised method works – there is still something odd about what this paper does in that they also have a *Guess the Winner* part of the experiment (which seems to have taken place after, and independently of, the subjective probability elicitation). They write *"We check the validity of the elicited probabilities by comparing them against a set of incentivized choices that we allow participants to make at the same moments when their subjective probabilities are elicited"* yet they do not discuss the potential usefulness of this data and do not compare these incentivised elicited subjective probabilities with the un-incentivised ones. Instead they use this data as described in the Appendix: *"Participants' winnings were summed up and paid out over the six "Guess the winner" games, and this total amount won was also used as the "Guess the winner" analysis variable, which is intended as an indicator of how well participants were able to predict the outcome of a race."* This is a sort of average measure for each subject, but surely the data could have been used in a similar way to the analysis using the un-incentivised data? I understand that the position of the cars at the various pit stops are different from those in the actual races, but there is information value in this data – and I therefore wonder why they have used it in the particular way that they did.

Indeed, continuing with this line of thought makes me wonder why the authors had the (incentivised) decision problem at all. It makes the experiment more difficult, and they do not exploit all the possibilities of analysing all the data together. True, they do do a probit analysis (reported in Table 2) of the decision at the final pit stop, but they do not look at the other pit stop decisions and nor at the amount-to-invest and amount-to-wager decisions. This makes me think that they could have fitted some preference functional – perhaps rank-dependent – subject by subject, using separately the objective probabilities (estimating the weighting function) and the subjective ones (possibly without a weighting function). The problem here is that the decision task facing the subjects is a difficult dynamic one, and has to be solved by some appropriate method. If (and this is a big 'if') the subjects solved the problem by backward induction, then we get one decision-rule. However, subjects could have been dynamically inconsistent, or solved the dynamic problem in some other way. So the subjective probability issue (which is the key concern of the paper) gets mixed up with how subjects solve a dynamic decision problem. Which again makes me wonder why the authors had the decision problem at all – it complicates the problem and they do not use all the data combined together. There is so much more that could be done with the data.

One thing that leaps out of Table 1 is that in over 99.9% of all races all subjects continued at least to the third pit stop – with very few opting out at earlier pit stops[1], though it is not clear how many dropped out at the third pit stop. I presume enough to construct Table 2 in which I think that the dependent variable takes the values 0 (dropping out) and 1 (not dropping out). But then I do not understand the *N* value of 3935 – which seems different from the 4299 quoted in Table 1. I should note that the hypothesis implicit in the probit analysis reported in Table 2 must rely either on risk-neutrality or constant attitude-to-risk across all subjects.

I would like to get some feel for the spread of subjective and objective probabilities. I presume from the subsequent analyses that these almost covered the entire spectrum from 0 to 1. Would a simple scatter of the one against the other be interesting and informative?

I would also like to make a rather important point about the econometrics. In all the specifications relating the subjective to the objective probabilities (equations (1), (2), (3) and (4)) there is an error term – which is always assumed to normal and homoscedastic. While I understand that this makes the econometrics easier, it is rather odd. First, since a normal variable is unbounded this implies that the subjective probabilities are too. Surely subjects were constrained to specifying a number between 0 and 1. Second, I would have thought that the variance would differ depending upon the objective probability. As that gets closer to 0 and 1 surely the variance would get smaller. I would suggest postulating that the subjective probabilities have a *beta distribution* with parameters α and β which are related to the expressions on the right-hand sides of equations (1), (2), (3) and (4) – which I denote by $\underline{p}^s_{it}$ in what follows – as follows: $\alpha = \underline{p}^s_{it}/(s-1)$ and $\beta = (1 - \underline{p}^s_{it})/(s-1)$, so that the mean of p^s_{it} is $\underline{p}^s_{it}$ while its variance is $\underline{p}^s_{it}(1-\underline{p}^s_{it})/s$. This has much nicer properties than a normal distribution and captures the point made above that the variance of p^s_{it} should get smaller as $\underline{p}^s_{it}$ approaches 0 or 1. I would seriously suggest re-computing many of the tables with this alternative stochastic specification. It may require some programming rather than the use of standard econometric packages.

Let me comment on the attitude to risk of the subjects. Clearly the attitude to risk varies across subjects. The authors do not try to measure this directly, and rely instead on the demographics (and other variables) to capture the variability. But I wonder about the reliability of this questionnaire data.

I do not like the references to brain activity – they seem superfluous, pointless and unnecessary.

All in all, I find this a difficult paper to assess, and find it hard to come up with a recommendation to the editors. As I said at the beginning it is a very complicated experiment, with much of it irrelevant to the task-in-hand, namely discovering the relationship between subjective probabilities and objective ones. The experiment generated a vast amount of data, only part of it which is used in a systematic way. The econometrics is suspicious.

My first best recommendation is to ask the authors to fit a structural model. But this would be difficult. If I have to make a realistic recommendation, then it would be for a *revise and resubmit* with the econometrics done differently, and the authors discussing the points raised above. I would be happy to look at a re-submission.

[1] With 239 participants each with 6 races and 3 pit stops, if all continue to the third pit stop we would see 239 times 6 times 3 which equals 4302 observations – but this was observed 4299 times. Perhaps I am misunderstanding this table. Incidentally what does the last row mean?