Collaborative ODI Research

Review by Professor Stephane Hess

Background

1. This review summarises my appraisal of the work carried out by PJM economics for the above project. I am an expert in choice modelling and stated preference survey design and am familiar with the background of the project as well as the state-of-practice in the field. I am also familiar with related previous studies producing valuation evidence in a water context. I am thus well positioned to provide this review.
2. I have had a number of meetings with the project team during which I provided feedback at intermediate stages of the work. This feedback has been taken on board in revisions made to the methodology and I commend the project team for their engagement with my feedback.

Survey work

3. In common with a majority of work on consumer valuations across different disciplines, including water research, the present project relied on the use of stated preference data, where respondents are faced with hypothetical scenarios that are used to elicit their preferences and/or monetary valuations.
4. In the context of capturing valuations from a large cross-section of the population in the context of rare events such as studied in the present project, the reliance on stated preference data is the only realistic option. It allows analysts to capture valuations from individuals who have not been exposed to such disruptions and in addition does so for a wide range of disruptions.
5. The present project made a substantial departure from previous studies in the survey approach.
    a. Past work in this area had made use of an approach that presented respondents with a choice between different services, each characterised by service levels (i.e. the risk of different disruptions) and a cost.
    b. By contrast, the *new* approach avoids the use of service levels and instead focusses on capturing the relative impact on households of different service disruptions occurring with certainty, and then monetises these through a separate compensation exercise.
6. The *existing* approach relies on respondents being able to adequately understand small risk levels, where academic evidence shows that this is often not the case. In addition, the resulting values may be affected by respondents being risk prone or risk seeking, and also by considering the risk of such disruptions happening to others, not just themselves. The new approach avoids the need for respondents to understand small risks, removes the potential for results to be affected by risk averseness or risk proneness, and also increases the likelihood of results relating to valuations of the disruption caused to households themselves, rather than wider impacts.

7. The *existing* approach allows analysts to capture the willingness-to-pay (WTP) of respondents to avoid given levels of risk of disruptions. In contrast, the *new* approach produces willingness-to-accept (WTA) measures, relating to the required level of compensation in case of disruption. Which of these measures is more appropriate depends on the application context.
8. The *new approach* has the potential disadvantage that the resulting valuations need to be obtained by combining the results from the impact stage with those from the valuation stage. This relies on fungibility between the two experiments, but reduces respondent burden in comparison with a joint approach. The combination of impacts and valuations in a single experiment was tested during a separate pilot following an earlier suggestion by this peer reviewer, and evidence was produced that the use of a dual experiment was preferable.
9. On balance, I am of the opinion that the advantages outweigh the disadvantages in moving from the *existing* to the *new* survey approach. I am satisfied with the arguments put forward by the project team during detailed discussions with me, as also summarised in their report. I also commend the project team on the extensive tests carried out to test the appropriateness of their approach.

Sampling, survey design and testing

10. I have no specific comments on the sampling. I feel that the best possible effort was used to obtain high-quality samples.
11. I commend the project team for engaging with the state-of-the-art in stated choice surveys by making use of efficient designs and using the results from the pilot to generate priors for improving the design.
12. The survey was well designed and administered. Care was used in the presentation to respondents, with detailed guidance provided in the survey.
13. I find the testing that was conducted to be of high quality, ensuring reliable data.
14. A careful weighting procedure was used to ensure that the results are representative for the target population. This applies to both the household and non-household samples.

Modelling work: SP1

15. The modelling work carried out for the analysis of the SP1 data is of high quality. It uses state-of-the-art Bayesian estimation of Mixed Logit models. The project team allowed for a fully flexible model specification, capturing correlation between the distributions for individual parameters. This goes beyond the state-of-practice.
16. Bayesian estimation involves, just like classical estimation, a number of subjective judgement calls. With Bayesian estimation, this relates in particular to the distributional assumptions and the decisions on the number of burn-in and post burn-in iterations, and chain thinning. Again, the decisions made are well argued, equate to best practice, and the project team is to be commended for the care taken in convergence testing.
17. The tests carried out on the data prior to the analysis as well as on the results (sensitivity analysis, subpopulation analysis and influence of experience, usage

and attitudes) are detailed and increase the robustness of the findings. Especially the validity tests for users are convincing.

Modelling work: SP2

18. The modelling work carried out for the analysis of the SP2 data is in line with best practice. I support the decisions made in terms of Q1 and Q2 responses, the use of a parametric model, and the choice of lognormal distributions. The decisions are not only supported by careful arguments and tests, but are also beneficial in terms of study outputs.
19. As with SP1, the testing of the data and results is of high quality.

Derivation and validation of values

20. The method used to combine the results from SP1 and SP2 is sound and is the most robust approach possible aside from joint estimation.
21. The discussion of the results is detailed and careful, and the comparison with the PR19 results provides is insightful. I believe that the reasons put forward for the differences in valuation are sensible.
22. The gap in valuations at the upper end is arguably surprisingly small. Going from an unexpected 24 hour water supply interruption to a two month emergency drought restriction would in general appear to entail a greater loss in utility than implied by the increase in valuation from £204 to £236. The same can be said for the two sewer flooding events.

Conclusion

23. I am of the opinion that the study was conducted in a very robust manner, and that the decisions taken to address key specification issues are justified and likely to have improved the robustness of results.

Stephane Hess

16 May 2023 (edited 20 July 2023)