11 March 2018

Dear Professor Whitehead,

I have now received reports from the original authors and three reviewers regarding your manuscript, "A Replication of Willingness-to-pay Estimates in 'An Adding Up Test on Contingent Valuations of River and Lake Quality' (Land Economics, 2015)" (*Manuscript Number 2092; Discussion Paper Number 2017-55*).

Based on the feedback I have received, and your efforts to address the responses of the reviewers in the revised manuscript you sent me, I am inviting you to submit a further revision.

Your replication produces tests for four approaches: The nonparametric *Turnbull* procedure and three parametric procedures (*Linear Logit: Mean WTP*, *Linear Logit: Mean WTP > 0*, and *Log Linear Logit: Median WTP*). In two of the four cases (*Turnbull*, *Linear Logit: Mean WTP > 0*) you reject the adding up test, while failing to reject in the other two cases (*Linear Logit: Mean WTP*, *Log Linear Logit: Median WTP*). But even in the two cases where you fail to reject, the numerical gap between the two WTP values are very large.

For *Linear Logit: Mean WTP*, the Whole and Sum of the Parts are 250 and 418, respectively. For the *Log Linear Logit: Median WTP*, they are 201 and 359. Those are fairly large differences, even if they are not statistically significant. And I don't have to remind you that failure to reject the null is not the same thing as accept.

So there are several additional things I would like you to do. First, I would like you to evaluate the four models and discuss the pros and cons of the respective approaches. In particular, is there any reason to prefer one of the models more than the others?

In this context, you need to address the points made by the original authors. Namely, that the estimated coefficients for the *Linear Logit: Mean WTP* model imply "that 35% of the population has a negative WTP for the whole program, and over half of the population (55%) has a negative WTP for the program's 2nd increment". Likewise, using the *Log Linear Logit* model, "the estimated parameters imply that 24% of the population has a WTP of more than $10,000 for the program, and that 7.3% has a WTP of more than $1 million". If the original authors are correct in these statements, these should be explicitly acknowledged in your revision and discussed in the context of whether we should believe the WTP numbers from these parametric models.

I would also like you to address the original authors' following statement: "…inadequate response to the cost prompts remains a fundamental issue that has not been investigated as extensively as inadequate response to scope, but is perhaps even more important". That should also be discussed in the context of how seriously we should take the results of these respective adding up tests.

Further, I think you can do better with respect to testing the adding up hypothesis. Currently, you simply test for zero difference between the Whole and the Sum of the Parts WTP estimates. The problem with this is that failure to reject does not mean that the differences are the same. To get at that, you should do an equivalence test. This not so common in

economics, but does get used in psychology. A good reference is: Daniël Lakens, "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses", *Social Psychological and Personality Science*, Vol 8, Issue 4 (2017): 355 – 362.

Finally, related to the above, it would be good if you brought in some discussion of statistical power in the context of your analysis.

A revised version of your manuscript that addresses the issues above will be reconsidered for publication. If you choose to resubmit a revision, be sure to include a detailed discussion of how you responded to each of the comments above.

Once again, thank you for submitting your manuscript to *Economics; The Open-Access, Open-Assessment E-Journal.* I look forward to receiving your revision.

Sincerely,

Bob Reed
Coeditor, Economics E-Journal