

Comments on:
A Replication of Willingness-to-pay Estimates in
'An Adding Up Test on Contingent Valuations of River and Lake Quality'
(Land Economics, 2015)

Overall, I think the paper needs to start with a discussion of adding up and why your analyses are relevant. The paper is very mechanistic with little intuition or conceptual guidance. It is written as a note to a peer with whom you have an ongoing discussion, but to the outsider they will likely ask why is this relevant and how do the empirical results address the relevant issue.

P. 2 – “One theoretical validity test is for whether the percentage of respondents who are willing to pay the cost declines as the cost increases. DMT’s data suffers from nonmonotonicity (i.e., the percentage does not always decrease as the bid increases) and flat portions of the bid curve.”

Flat portions of the response function do not violate theoretical validity nor monotonicity, but you seem to accidentally imply that they do in your statement.

P. 2 – “Another reason for the replication is that the cost range does not cover the entire WTP distribution. In other words, the highest cost amount does not cause the percentage of yes responses to fall to zero.”

This is a two-sided issue, either the full “value” range is not covered, or some people will never say no to any bid. You do not have enough information to say which or if both are occurring in the data. Also, putting a bid in the tail of the distribution contributes little to estimation.

Section 2 – I suggest briefly reviewing what DMT valued and the definition of each of the scenarios, so the reader does not need to go back to the original paper to understand these treatments. A fundamental issue in adding up is whether it is accepted that the parts should in fact sum to the whole.

Table 1 – Identify the monotonicity violations with bolding or shading to make them apparent to the reader given the amount of data in the table.

Table 1 – An issue that you do not mention when discussing monotonicity violations and flat portions of the response functions is that the number of observations at each bid point is small; in some cases, not large enough for the law of large numbers to apply. This is an issue I am not aware of being discussed in the literature. Is there a need to think about sample sizes by bid amount, not just by the total sample? This has implications for choice experiments and experimental economics as well where bid amount effects are rarely considered.

Figure 1 – Was the DMT replication consequential? If not, then there is no reason for subjects to take the bids seriously and this could lead to the response pattern you show. The Chapman et al. study was consequential.

Carson, R.T. and Groves, T., 2007. Incentive and informational properties of preference questions. *Environmental and resource economics*, 37(1), pp.181-210.

Carson, R.T., Groves, T. and List, J.A., 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, 1(1/2), pp.171-207.

P. 5 – “I combine the data from the subsamples and estimate linear and log linear parametric dichotomous choice models as recommended by Boyle (2017)”

I do not see this recommendation in Boyle, but see no problem with this approach. However, when pooling data, you do this because correlation in responses across treatments yields estimation efficiency. You may want to think about the structure of the correlated error terms.

Table 5 – Identify that the t-stats are for whether the estimates are different from zero.

P. 6 – “The parametric WTP estimates are significantly (economically) different than the nonparametric estimates.”

I do not think you want to say “significantly” because you did not do statistical tests, and such test would be questionable because these are not data from independent samples. They are “economically relevant”.

P. 6 – “The null hypothesis of equality between WTP for the whole scenario and WTP for the sum of the parts cannot be rejected in two of the three adding up tests.”

Specify these hypotheses, identify the tests conducted and report the test statistics. While WTP is the policy-relevant statistic, why not test equivalence of the underlying responses functions (constant + slope)? An important issue you do not address is the fat tails on the distributions. These tails add variance to the estimation that may lead to failure to reject the null of no difference with the parametric estimator. Thus, one could argue that you used the wrong specification and your result is an artifact of your choice of a functional form that does not apply to the data.

P. 6 – “When I apply the same post-stratification weights, scaled to equal the sample size of $n=980$, to the models in Table 4 and estimate WTP as in Table 5, none of the three sets of parametric WTP estimates supports rejection of the null hypothesis of equality between WTP for the whole and the sum of the parts.”

This is true, but it only addresses the whole sample, not samples for individual bid amounts and their inherent response patterns. I am not sure your robustness analysis address fundamental estimation issues and contributes much to the paper.

1st para. of Conclusions – This reiteration of results is not needed in such a short paper.

P. 7 – I concur with the first two paragraphs and think your points can be stated more clearly. I find the last paragraph problematic for the issues raised above and this paragraph takes away from the impact of the content from the two preceding paragraphs.