

A Reply to John Whitehead's Replication Paper

by

William Desvousges^{a*}, Kristy Mathews^b, Kenneth Train^c

^aPresident, W.H. Desvousges & Associates, P.O. Box 99203, Raleigh, NC 27624

^bOwner and Principal, 104 McWaine Lane, Cary, NC 27513

^cAdjunct Professor, Department of Economics, University of California, Berkeley, 530 Evans Hall #3880, Berkeley, CA 94720-3880

*Corresponding author, william.desvousges@whdesvousgesassociates.com, 919-413-6225.

Whitehead uses two parametric estimators that extrapolate the distribution of willingness-to-pay (WTP) beyond the cost prompts that were administered in the survey. However, researchers have previously noted (see, e.g., Desvousges et al. 1992; Haab and McConnell 2002) that parametric approaches to CV data can be unreliable and unstable, giving implausible and counter-intuitive results. Whitehead's estimators exhibit these problems.

He first applies a logit model with cost as an explanatory variable, under the assumption that WTP follows a logistic distribution. The estimated coefficients of this 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. Adding-up tests based on such a large share of incorrect values for WTP are not informative.

Whitehead also uses a log-logit model under the assumption that the log of WTP follows a logistic distribution. This specification avoids the problem of negative WTPs, but introduces problems on the other side of the distribution. The estimated parameters of this model 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. The estimated cost coefficient is so small that the tail of the estimated distribution extends indefinitely, such that the mean WTP is estimated to be infinite. An adding-up test cannot be applied when the estimated mean WTP is infinite. However, Whitehead uses the estimated median WTPs instead of the means to conduct his adding-up test on this model. But the use of

medians masks the problem rather than addressing or solving it. Clearly, this estimated distribution is so unreasonable that any testing based on it is uninformative.

The problems with Whitehead's models are demonstrated directly in the implementation of the tests. For his linear logit model, the sum of the point estimates of the mean WTP for the parts is 67% greater than the point estimate of the mean WTP for the Whole. For the log-linear logit, the sum of the point estimates of the median WTPs for the parts is 79% greater than the point estimate of the median WTP for the Whole. When point estimates are so different and yet equality cannot be rejected, the quality of the model on which the test is based comes into question. A poor model with large standard errors is more likely to pass an equality test, like the adding-up test, than a better model with smaller standard errors. Careful evaluation of the plausibility of any parametric model results should precede any attempt at hypothesis testing.

Whitehead does, however, raise a very important issue by reminding us of the relative flatness of CV response curves and the difficulty that this creates for estimation of mean WTP. The CV debate has seen a lot of time and research funds spent on the issue of inadequate response to scope. But CV responses also evidence inadequate response to the cost prompts, and this issue is as important as the issue of scope. As Whitehead noted, Parsons and Myers (2016) recently addressed this issue by examining the typical phenomenon of "fat tails" in CV responses, by which the share of "yes" votes does not seem to approach zero as the cost prompt is raised. They reviewed numerous CV studies

and found that, typically, the share of “yes” votes was still fairly high at the highest cost prompt that was used in the study.

These fat tails make parametric methods unreliable and often unreasonable, which is one reason the profession moved to the nonparametric ABERS estimator (Ayer et al., 1955) for natural resource damage assessment (NRDA). But Parsons and Myers point out that ABERS does not actually address the issue of fat tails. Their review of CV studies that used ABERS found that each study’s estimated WTP was largely determined by the study’s highest cost prompt, because the “yes” shares were so high at the highest cost prompt. They tried to determine whether ABERS could become reliable by raising the highest cost prompt. Using a CV survey about protection of an endangered shorebird, they attempted to find the cost prompt at which the “yes” share approached zero, and they were not able to find one. They raised the cost prompt as high as \$10,000 and found that still 23% of respondents said that they would vote in favor of the program at that cost. Because of these fat tails, essentially any estimate of WTP can be obtained with ABERS through the researcher’s selection of the highest cost prompt. The ABERS estimator does not solve the problem of fat tails: its results are highly dependent on the researchers’ decision of how far out the tail to go.

The issue of inadequate response to cost arises on the other side of the distribution too: the share of “yes” votes is typically far below 100% at the lowest cost prompt. As a result, the cost prompts in a typical CV study cover only a small share of the distribution of WTP. For example, in the original Chapman et al. (2009) study that we used as the

basis for our analysis of adding-up, the “yes” share was 82% at the lowest cost prompt and 34% at the highest cost prompt, which means that the cost prompts covered only 48% of the density of WTP.¹ With so little coverage of the distribution, methods to estimate the mean WTP become unstable with respect to different parametric specifications and prone to giving implausible results.

Whitehead’s paper demonstrates the importance of the long-recognized difficulties with CV responses for estimating mean WTP. Although these difficulties led to the adoption of the ABERS estimator for NRDA, it does not actually solve them. Instead, 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.

Whitehead’s paper becomes a call to CV practitioners to acknowledge, and address if possible, the problem of estimating mean WTP from CV responses.

REFERENCES

Ayer, Miriam, H. D. Brunk, G.M. Ewing, W.T. Reid, and Edward Silverman. 1955. “An Empirical Distribution Function for Sampling with Incomplete Information.” *The Annals of Mathematical Statistics* 26(4): 641–647 (December).

Chapman, David J., Richard C. Bishop, W. Michael Hanemann, Barbara J. Kanninen, Jon A. Krosnick, Edward R. Morey, and Roger Tourangeau. 2009. *Natural Resource Damages Associated with Aesthetic and Ecosystem Injuries to Oklahoma’s Illinois River System and Tenkiller Lake*. Expert Report for State of Oklahoma,

¹ 82% - 34% = 48%.

Volume I, available at <https://pcl.uscourts.gov/search> (*Oklahoma v. Tyson Foods Inc.*, No. 4:05-cv-329 (N.D. Okla. Feb. 13, 2009), Docket No. 1853-4, exhibit D).

The public-use copy from this site is also available at

<http://elsa.berkeley.edu/~train/chapman.pdf>

Desvousges, W.H., F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson, and K.N.

Wilson. 1992. *Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy*. Monograph 92 1, Prepared for Exxon Company, USA. Research Triangle Park, NC: Research Triangle Institute.

Haab, Timothy C. and Kenneth E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Edward Elgar Publishing.

Parsons, George and Kelley Myers. 2016. "Fat Tails and Truncated Bids in Contingent Valuation: An Application to an Endangered Shorebird Species." *Ecological Economics* (129): 210-219.