Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2017-55 | August 21, 2017 | http://www.economics-ejournal.org/economics/discussionpapers/2017-55

A replication of willingness-to-pay estimates in 'An adding up test on contingent valuations of river and lake quality' (Land Economics, 2015)

John C. Whitehead

Abstract
Desvousges, Mathews and Train (2015) find that their contingent valuation method (CVM) survey data does not pass the adding up test using a nonparametric estimate of mean willingness-to-pay. Their data suffers from non-monotocity, flat bid curve and fat tails problems, each of which can cause willingness-to-pay estimates to be sensitive to the approach chosen to measure the central tendency. Using additional parametric approaches that are standard in the literature, I find that willingness to pay for the whole is not statistically different from the sum of the parts in two of three additional estimates. In additional robustness checks, all six of the additional tests find that the WTP estimates do not reject the adding up hypothesis. The negative result in Desvousges, Mathews and Train (2015) is not robust to these alternative approaches to willingness-to-pay estimation.

(Submitted as Replication Study)

JEL Q51
Keywords Contingent valuation; adding up test; willingness-to-pay

Authors
John C. Whitehead, Department of Economics, Appalachian State University, Boone, NC, USA, whiteheadjc@appstate.edu

Citation John C. Whitehead (2017). A replication of willingness-to-pay estimates in 'An adding up test on contingent valuations of river and lake quality' (Land Economics, 2015). Economics Discussion Papers, No 2017-55, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2017-55

Received January 18, 2017 **Accepted as Economics Discussion Paper** August 10, 2017
Published August 21, 2017

1 Introduction

The contingent valuation method (CVM) is a stated preference survey approach to the valuation of public goods (Mitchell and Carson 1989, Haab and Whitehead 2015). Desvousges, Mathews and Train (2012), in a critique of the Chapman et al. (2009) natural resource damage assessment, argue that CVM studies must pass the "adding up test" to demonstrate "adequate" responsiveness to scope (see also Whitehead 2016, Chapman et al. 2016 and Desvousges, Mathews and Train 2016). Desvousges, Mathews and Train (DMT, 2015) field the Chapman et al. (2009) survey with new sample data and additional scenarios. DMT argue that willingness-to-pay (WTP) for the whole should be equal to willingness to pay for the sum of four parts (the first, second, third and fourth scenarios). DMT find that "The sum of the four increments … is about three times as large as the value of the whole" (p. 566). In this replication I examine DMT's conclusion using alternative parametric approaches for estimating the central tendency of WTP.[1]

This replication is appropriate for several reasons. Dichotomous choice contingent valuation questions propose a cost to respondents who then indicate whether or not they are willing to pay the cost. 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 non-monotonicity (i.e., the percentage does not always decrease as the bid increases) and flat portions of the bid curve. 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 "fat tails" problem is pervasive in CVM data (Parsons and Myers 2016) and causes WTP to be sensitive to the estimation approach.

Following Chapman et al. (2009), DMT choose the ABERS nonparametric estimator for willingness to pay (Ayer et al. 1955). Chapman et al. (2009) describe the ABERS estimator as producing a lower bound WTP estimate. The ABERS estimator is equivalent to the more familiar Turnbull nonparametric lower bound WTP estimator (Haab and McConnell 1997, Carson and Hanemann 2005, Boyle 2017). Both nonparametric WTP estimation approaches obscure data quality problems. When data is non-monotonic, both ABERS and Turnbull approaches smooth non-monotonic bid curves by pooling percentages of those willing to pay across cost amounts and ignore validity problems associated with flat portions of the bid curve. Both the ABERS and Turnbull estimates truncate the WTP distribution at the highest bid, ignoring the tail of the WTP distribution.

In the remainder of this paper I replicate the ABERS willingness-to-pay estimates with the Turnbull and reproduce DMT's negative result on the adding-up test. In section three I present two parametric models of WTP that lead to three additional WTP estimates for each scenario. One of these estimates supports DMT's negative adding up test result but two fail to reproduce the negative result. In the fourth section, I conduct the same analysis with post-stratification weights and a more reliable subsample of (complete case) data. All six of these adding up tests fail to reproduce the negative adding up result in DMT (2015). In the conclusions I offer recommendations on future CVM studies, one of these is to conduct sensitivity analysis over WTP estimation approaches when CVM data is "difficult."

[1] In a separate comment I argue that the adding up test survey design is flawed (Whitehead 2017).

2 WTP Replication

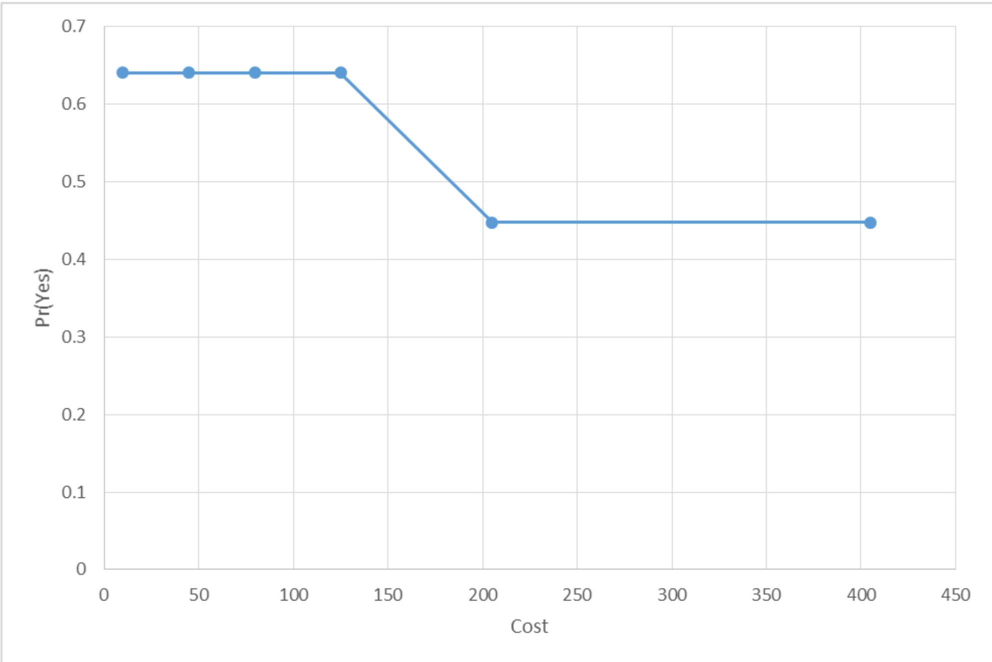
The data from DMT is presented in Table 1. Each of the scenarios exhibits non-monotonicity in at least one of the five cost increases. For example, in the whole scenario the percentage yes is 61 at \$45 and 69 at \$80. The whole, first and fourth scenarios exhibit non-monotonicity in the cost increase from \$205 to \$405.

Table 1. Dichotomous Choice CVM Data (DMT 2015)

	Whole			First			Second			Third			Fourth		
Cost	Yes	N	%Yes	Yes	N	%Yes	Yes	N	%Yes	Yes	N	%Yes	Yes	N	%Yes
10	17	25	68	38	51	75	12	24	50	24	29	81	24	33	73
45	20	33	61	28	48	58	12	32	38	13	27	48	11	25	44
80	18	26	69	31	48	65	7	24	29	10	31	32	24	37	65
125	14	28	50	27	47	57	12	28	43	6	26	23	20	32	63
205	13	29	45	21	54	39	6	25	24	11	27	41	10	28	36
405	14	31	45	18	45	40	4	26	15	12	34	35	11	27	41
Total	96	172	56	163	293	56	53	159	33	76	174	44	100	182	55

Even when the yes responses are monotonically decreasing in the cost amount, the slope is not different from zero in large portions of the bid curves. For example, the whole and second scenarios are characterized by two flat portions of the bid curve. A stylized example is illustrated in Figure 1 where the percentage of yes responses is constant over the lower range of cost amounts (\$10 to \$125), is downward sloping from \$125 to \$205 and flat from \$205 to \$405.

Figure 1. Bid curve with two flat portions



For the whole scenario, the slope of the bid curve over the entire range of cost amounts (\$10 to \$405) is downward sloping with $b = -.00058$ ($t = -2.09$, n = 172) estimated with a linear probability model ($\Pr(Yes) = a + b \times Cost$). The slopes over the lower ($10 to $80) and upper ($125 to $405) ranges of cost amounts are flat with $b = 0.0019$ ($t = 0.10$, n = 84) and $b = -0.00013$ ($t = -0.29$, n = 88), respectively. Similarly, in the second scenario the slope of the bid curve over the entire range of cost amounts ($10 to $405) is downward sloping with $b = -.00074$ ($t = -2.63$, n = 159). The slopes over the lower ($10 to $125) and upper ($205 to $405) ranges of cost amounts are flat with $b = -0.00056$ ($t = -0.49$, n = 109) and $b = -0.00043$ ($t = -0.76$, n = 51), respectively.

With non-monotonic data, nonparametric WTP estimators require pooling of yes responses across cost amounts until monotonicity is achieved. When the probabilities for two pooled costs are higher than the next lowest cost the pooling continues until the bid curve is non-monotonically non-increasing in the cost amount. The pooled dichotomous choice data are presented in Table 2.

Table 2. Monotonically Non-increasing Probability of a Yes Response

	%Yes				
Cost	Whole	First	Second	Third	Fourth
10	68	75	50	83	73
45	64	61	38	48	59
80	64	61	37	33	59
125	50	57	37	33	59
205	45	39	24	33	38
405	45	39	15	33	38

The lower bound Turnbull WTP estimate is the step function formed by the data in Table 2 (Haab and McConnell 1997, 2002). The Turnbull WTP estimates are presented in Table 3 with standard errors (SE) computed as in Haab and McConnell (1997, 2002), a common approach found in the CVM literature (see e.g., Egan, Corrigan and Dwyer 2015). The Turnbull WTP estimates are equal to the ABERS WTP estimates presented by DMT when rounded.

Table 3. Nonparametric Willingness to Pay Estimates

	DMT (2015)		Replication	
	ABERS	SE	Turnbull	SE
Whole	200	17.71	200.38	19.65
First	187	12.31	186.63	15.03
Second	97	13.73	97.33	18.16
Third	144	15.34	144.11	22.69
Fourth	181	18.69	181.47	23.66

The null hypothesis consistent with the adding up test as discussed by DMT is $HO\colon WTP_w = \sum_{i=1}^{4} WTP_i$, where w is the whole scenario and $i = 1, 2, 3, 4$ indicates the first, second, third and fourth scenarios. The alternative hypothesis is the inequality. With the Turnbull estimates $\sum_{i=1}^{4} WTP_i = 610$ which is $409 greater than WTP_w. The larger Haab and McConnell standard errors will favor the null hypothesis of the adding up test. Nevertheless, with the standard error for the sum of the four parts constructed as the square root of the sum of the variances of the four parts (SE = 45) (Haab and McConnell 2002)[2], the WTP estimates fail the adding up test, replicating the result in DMT (2015).

3 Parametric Estimates of WTP

In order to investigate the robustness of DMT's results, I combine the data from the sub-samples and estimate linear and log linear parametric dichotomous choice models as recommended by Boyle (2017): $\ln(\Pr(Yes)/(1-\Pr(Yes)) = a + b \times Cost$ and $\ln(\Pr(Yes)/(1-\Pr(Yes)) = a + b \times lnCost$ These models are specified so that each scenario (whole, first, second, third and fourth) has its own constant and its own cost variable. The models are estimated using LIMDEP version 10 (http://www.limdep.com).

In each of the models the slope coefficients (b) are statistically different from zero (Table 4). In the linear logit model the constants for the whole, first, and fourth scenarios are statistically different from zero. In the log linear logit all constants except in the second scenario are statistically different from zero. The log linear model provides a better statistical fit than the linear logit.

Table 4. Dichotomous Choice Probability Models

	Linear Logit			Log Linear Logit		
Constant (a)	Coefficient	SE	t-stat	Coefficient	SE	t-stat
Whole	0.594	0.235	2.53	1.58	0.653	2.42
First	0.726	0.182	4.00	2.11	0.503	4.19
Second	-0.190	0.249	-0.76	0.960	0.664	1.45
Third	0.145	0.229	0.64	2.19	0.644	3.39
Fourth	0.610	0.225	2.70	1.73	0.617	2.81
Slope (b)						
Whole	-0.0023	0.0012	-2.05	-0.298	0.141	-2.14
First	-0.0035	0.0010	-3.65	-0.422	0.108	-3.90
Second	-0.0039	0.0015	-2.51	-0378	0.154	-2.54
Third	-0.0027	0.0012	-2.29	-0.549	0.146	-3.91
Fourth	-0.0030	0.0011	-2.45	-0.347	0.136	-2.60
χ^2		66.08			80.67	
McFadden R^2		0.05			0.06	
Sample size		980			980	

The parametric willingness to pay estimates are presented in Table 5. Mean (and median) WTP from the linear logit, which allows negative WTP, is the negative ratio of the constant and the slope: $WTP = -a/b$ (Hanemann 1984). Estimating WTP only over the positive portion of the distribution from the linear logit uses the formula: $WTP = \left(\frac{-1}{b}\right)\ln(1 + \exp(a))$ (Hanemann 1989). Median WTP from the log linear logit is the exponential of the negative ratio of the

[2] DMT "applied the bootstrap method to simulate the sampling distribution of the difference between the mean WTP for the whole and the sum of the mean WTP from the four increments."

constant and slope: $WTP = \exp\left(-\frac{a}{b}\right)$. Mean WTP from the log linear model is undefined when $-\frac{1}{b} > 1$ (Haab and McConnell 2002) as in these models. Standard errors are estimated with the Delta Method (Cameron 1991).

Table 5. Willingness-to-pay Estimates

	Linear Logit						Log Linear Logit		
	Mean WTP			Mean WTP > 0			Median WTP > 0		
	WTP	SE	t-stat	WTP	SE	t-stat	WTP	SE	t-stat
Whole	250	81	3.09	434	171	2.56	201	126	1.59
First	208	39	5.34	321	66	4.87	149	46	3.21
Second	-49	80	-0.62	156	46	3.42	13	11	1.20
Third	54	69	0.78	285	96	2.96	54	17	3.17
Fourth	205	58	3.53	352	112	3.13	147	71	2.07
Sum of Parts	418	127	3.29	1114	168	6.63	359	92	3.90

The parametric WTP estimates are significantly (economically) different than the nonparametric estimates. Considering the whole scenario, the WTP estimates are 25%, 117% and 0.5% larger than the Turnbull estimates. The similarity between the mean Turnbull and the median WTP from the log-linear model may be only coincidence since the two estimates are based on different measures of central tendency. Considering the sum of the parts, the WTP estimates are -31% smaller, 83% larger and -41% smaller than the Turnbull estimates.

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. The linear logit that allows for negative mean WTP estimates yields a difference of $168 that is not statistically different from zero as the 95% confidence intervals overlap. These WTP estimates pass the adding up test. In the linear logit with the mean WTP constrained to be positive the difference between the whole and the sum of the parts is $680. The upper limit on the 95% confidence interval for the whole scenario is 766. The lower limit on the 95% confidence interval for the WTP for the sum of the parts is 785. These WTP estimates fail to pass the adding up test. The log linear logit produces a difference of $187 in median WTP that is not statistically different from zero. The median WTP estimates pass the adding up test.

4 Further Robustness Checks

DMT report that they conducted sensitivity analysis using post-stratification weights and present regression results with a sample smaller than that used for the mean WTP estimation. In this section I conduct the parametric analysis with these weights and this alternative sample. DMT report that the post-stratification weights do not change the nonparametric results. 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.

However, these results are complicated by incorrect signs and statistically insignificant WTP estimates for the problematic whole and second scenarios. The weighted models produce incorrect signs on the constant and slope in the second scenario (see Appendix, Figure 1). The incorrect signs lead to a positive weighted WTP estimate of $346 (SE=49) in the second scenario when WTP is estimated over the entire range. The weighted WTP estimate is -$34 (SE=15) when it is estimated only over the positive range. But, both of these WTP estimates are nonsensical given the positive relationship between cost and the probability of a yes response.

Considering the whole scenario, the weighted WTP is $1154 (SE=1289) when estimated only over the non-negative range and the sum of the weighted WTP parts is $811 (SE = 212) (see Appendix, Figure 1). The statistically insignificant weighted mean WTP for the whole scenario leads to wide confidence intervals for which it is difficult to reject the null hypothesis of equality between WTP for the whole and the sum of the parts.

DMT (2015) conduct their nonparametric WTP estimation with a full sample of n=980. Yet, they conduct regression analysis with a sample of n=950 in order to estimate income effects. Close examination of the data reveals that there are only 936 cases that do not suffer from item nonresponse. Forty-three cases have missing income values for which 14 unconditional means of the income variable are imputed for the n=950 regression analysis. There are 30 cases with item nonresponse in the age variable. These 30 cases are dropped for the n=950 regression analysis in DMT (2015). There is one missing age value that occurs with a nonmissing income value so the total number of cases with missing age and/or income values is 44 (see Appendix, Figure 2).

The percentage of yes responses for the 44 respondents who did not answer the age and/or income questions, 66% (n=44), is higher than for the complete case sample, 49% (n=936). Since it appears that this subsample is different than the complete case sample we re-estimate the models in Tables 4 and 5 discarding those who did not answer the age and/or income question. We find that all three of the adding up tests fail to reject the null hypothesis of equality between WTP for the whole and the sum of the parts. For example, the linear logit model with mean WTP estimated over the positive range is $445 (SE = 193) in the whole scenario and the sum of the WTP parts is $1080 (SE = 174) (see Figure 3 in the Appendix). The 95% confidence intervals for these estimates overlap.

Examination of the income effects estimated by DMT is beyond the scope of this paper. Nevertheless, it is worth mentioning that the unweighted models with the cost coefficient constrained to be equal across scenarios produces statistically insignificant income effects as in DMT (2015). Applying the post-stratification weights and allowing cost amounts to vary over the scenarios, as is statistically appropriate, leads to statistically significant income effects in the n=980 (with age imputed at the mean), n=950 and n=936 samples. These results, which suggest DMT (2015) may use an inappropriate income coefficient for their simulations, are available upon request.

5 Conclusions

Desvousges, Mathews and Train's (2015) dichotomous choice CVM data leads to WTP estimates that fail to reject the null adding up hypothesis test with two of three alternative parametric estimates of WTP. In addition, the weighted WTP estimates fail to reject the null hypothesis of equality between WTP for the whole and the sum of parts with all three parametric estimates. And using a subsample of data discarding respondents who do not answer the age and income questions, the WTP estimates fail to reject the null hypothesis with all three parametric estimates. DMT's results are not robust to alternative, but standard, parametric approaches to estimating WTP. The failure to replicate DMT's results with the parametric models is due to a host of data quality problems: non-monotonicity, fat tails and flat portions over wide ranges of the bid function. Each of these problems leads to high variability in mean WTP across estimation approach and larger standard errors than those associated with nonparametric estimators.

The data quality problems are particularly apparent in the whole and second scenarios which are the versions of the survey developed by Chapman et al. (2009). Considering this, DMT (2015), who use a relatively inexpensive, small non-probability sample and an online survey, fail to replicate the Chapman et al. (2009) study. Chapman et al. (2009) use in-person interviews with a large probability sample as recommended by Arrow et al. (1993). Many of the problems in the DMT (2015) data may be due to the lack of a large research budget. Researchers who are tempted to use these inexpensive panels with online surveys and small samples should do so with caution.

Future studies should attempt to address these problems with larger subsamples of data. This can be achieved in three ways. The most costly, of course, is simply by increasing the overall sample size. Holding cost constant, researchers could reduce the number of cost amounts used in the experimental design or reduce the number of experimental scenarios presented. For example, DMT could have implemented their adding up test with three (instead of five) separate scenarios as they describe in Desvousges, Mathews and Train (2012) (see also Whitehead 2017).

The nonparametric WTP estimate used by DMT is appropriate for natural resource damage assessment where a lower bound estimate is desired as in Chapman et al. (2009). The nonparametric WTP estimate, when considered in isolation from other WTP estimates, is less appropriate for validity testing of the CVM since it may minimize differences in WTP across samples. This, by construction, results in a bias against finding economically and/or statistically significant differences in WTP across survey treatments. With what Haab and McConnell (2002) call “difficult data,” the entire range of nonparametric and parametric WTP estimates should be examined for validity testing and benefit-cost analysis.

Acknowledgements The author thanks Bill Desvousges for providing the data.

References

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, Howard Schuman, Report of the NOAA Panel on Contingent Valuation, January 11, 1993. http://www.darp.noaa.gov/pdf/cvblue.pdf.

Ayer, Miriam, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman. 1955. “An Empirical Distribution Function for Sampling with Incomplete Information.” Annals of Mathematical Statistics 26 (4): 641–47. http://projecteuclid.org/euclid.aoms/1177728423.

Boyle, Kevin J. "Contingent valuation in practice." In A primer on nonmarket valuation, pp. 83-131. Springer Netherlands, 2017. http://www.springer.com/us/book/9789400771031

Cameron, Trudy Ann. "Interval estimates of non-market resource values from referendum contingent valuation surveys." Land Economics 67, no. 4 (1991): 413-421. http://www.jstor.org/stable/3146548?seq=1#page_scan_tab_contents

Carson, Richard T., and W. Michael Hanemann. "Contingent valuation." Handbook of environmental economics 2 (2005): 821-936. http://www.sciencedirect.com/science/article/pii/S1574009905020176

Chapman, David, Richard Bishop, Michael Hanemann, Barbara Kanninen, Jon Krosnick, Edward Morey and Roger Tourangeau. 2009. Natural Resource Damages Associated with Aesthetic and Ecosystem Injuries to Oklahoma’s Illinois River System and Tenkiller Lake.

Chapman, David J., Richard C. Bishop, W. Michael Hanemann, Barbara J. Kanninen, Jon A. Krosnick, Edward R. Morey, and Roger Tourangeau. "On the adequacy of scope test results: Comments on Desvousges, Mathews, and Train." Ecological Economics 130 (2016): 356-360. http://www.sciencedirect.com/science/article/pii/S0921800916306139.

Desvousges, William, Kristy Mathews, and Kenneth Train. "Adequate responsiveness to scope in contingent valuation." Ecological Economics 84 (2012): 121-128. http://www.sciencedirect.com/science/article/pii/S0921800912003813.

Desvousges, William, Kristy Mathews, and Kenneth Train. “An Adding Up Test on Contingent Valuations of River and Lake Quality.” Land Economics 91(2015): 556-571. http://le.uwpress.org/content/91/3/556.refs.

Desvousges, William, Kristy Mathews, and Kenneth Train. “Reply to ‘On the adequacy of scope test results: Comments on Desvousges, Mathews, and Train’,” Ecological Economics 130 (2016): 361–362. http://www.sciencedirect.com/science/article/pii/S0921800916306139.

Egan, Kevin J., Jay R. Corrigan, and Daryl F. Dwyer. "Three reasons to use annual payments in contingent valuation surveys: Convergent validity, discount rates, and mental accounting." Journal of Environmental Economics and Management 72 (2015): 123-136. http://www.sciencedirect.com/science/article/pii/S0095069615000443

Haab, Timothy C., and Kenneth E. McConnell. "Referendum models and negative willingness to pay: alternative solutions." Journal of Environmental Economics and Management 32, no. 2 (1997): 251-270. http://www.sciencedirect.com/science/article/pii/S0095069696909687.

Haab, Timothy C., and Kenneth E. McConnell. Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation. Edward Elgar Publishing, 2002. https://www.elgaronline.com/view/9781840647044.xml.

Hanemann, W. Michael. "Welfare evaluations in contingent valuation experiments with discrete responses." American Journal of Agricultural Economics 66, no. 3 (1984): 332-341. http://ajae.oxfordjournals.org/content/66/3/332.short.

Hanemann, W. Michael. "Welfare evaluations in contingent valuation experiments with discrete response data: reply." American Journal of Agricultural Economics 71, no. 4 (1989): 1057-1061. https://academic.oup.com/ajae/article-lookup/doi/10.2307/1242685

Parsons, George R., and Kelley Myers. "Fat tails and truncated bids in contingent valuation: An application to an endangered shorebird species." Ecological Economics 129 (2016): 210-219. http://www.sciencedirect.com/science/article/pii/S0921800915301567.

Whitehead, John C. "Plausible responsiveness to scope in contingent valuation." Ecological Economics 128 (2016): 17-22. http://www.sciencedirect.com/science/article/pii/S0921800916302890.

Whitehead, John C. "A Comment on 'An Adding Up Test on Contingent Valuations of River and Lake Quality.'" No 17-01, Working Papers from Department of Economics, Appalachian State University. http://econpapers.repec.org/paper/aplwpaper/17-01.htm.

Whitehead, John C., and Haab Timothy C. (2013) Contingent Valuation Method. In: Shogren, J.F., (ed.) Encyclopedia of Energy, Natural Resource, and Environmental Economics, Vol. 3, pp. 334-341 Amsterdam: Elsevier. http://www.sciencedirect.com/science/article/pii/B9780123750679000048.

Appendix. Additional LIMDEP output referred to in the text

Figure 1. Weighted linear logit model and positive constrained WTP estimates with post-stratification weights (n=980)

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                VOTE
Weighting variable               WT980
Log likelihood function      -609.10542
Restricted log likelihood    -676.28727
Chi squared [   9](P= .000)   134.36370
Significance level              .00000
McFadden Pseudo R-squared     .0993392
Estimation based on N =    980, K =  10
Inf.Cr.AIC  =   1238.2 AIC/N =    1.263
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
    VOTE|  Coefficient     Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
   WHOLE|     .16083        .18483      .87  .3842     -.20143     .52309
   FIRST|    1.08449***     .23223     4.67  .0000      .62932    1.53966
  SECOND|   -1.69575***     .25529    -6.64  .0000    -2.19610   -1.19540
   THIRD|     .24915        .27176      .92  .3593     -.28350     .78180
  FOURTH|     .79497***     .25280     3.14  .0017      .29950    1.29044
 AMOUNTW|    -.00067        .00081     -.83  .4070     -.00226     .00092
 AMOUNT1|    -.00524***     .00127    -4.13  .0000     -.00773    -.00275
 AMOUNT2|     .00490***     .00110     4.44  .0000      .00274     .00707
 AMOUNT3|    -.00579***     .00136    -4.26  .0000     -.00846    -.00312
 AMOUNT4|    -.00265*       .00150    -1.77  .0768     -.00559     .00029
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 06, 2017 at 00:16:05 PM
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors for nonlinear functions and
joint test of nonlinear restrictions.
Wald Statistic                =     80.00000
Prob. from Chi-squared[ 6] =         .00000
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
WaldFcns|   Function       Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
    WTPW|    1154.56      1288.863      .90  .3704    -1371.57    3680.68
    WTP1|    262.460***   43.98971     5.97  .0000     176.242    348.679
    WTP2|   -34.3673**    14.94890    -2.30  .0215    -63.6666    -5.0680
    WTP3|    142.540***   21.53797     6.62  .0000     100.326    184.753
    WTP4|    440.527**    205.7960     2.14  .0323      37.175    843.880
SUMPARTS|    811.160***   212.0718     3.82  .0001     395.507   1226.813
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 06, 2017 at 00:16:09 PM
```

Figure 2. Missing Age and Income values

Listing of current sample -----------------

Line	Observation	AGE	INC
1	41	41	43.19370
2	62	Missing	22.50000
3	109	58	43.19370
4	166	Missing	43.19370
5	168	46	43.19370
6	169	Missing	43.19370
7	170	Missing	43.19370
8	171	Missing	43.19370
9	172	Missing	43.19370
10	229	56	43.19370
11	280	56	43.19370
12	303	32	43.19370
13	325	Missing	43.19370
14	326	Missing	43.19370
15	327	Missing	43.19370
16	328	Missing	43.19370
17	329	Missing	43.19370
18	330	Missing	43.19370
19	331	Missing	43.19370
20	413	56	43.19370
21	495	Missing	43.19370
22	496	Missing	43.19370
23	497	Missing	43.19370
24	498	Missing	43.19370
25	499	Missing	43.19370
26	500	24	43.19370
27	501	Missing	43.19370
28	502	Missing	43.19370
29	503	Missing	43.19370
30	504	Missing	43.19370
31	505	Missing	43.19370
32	538	61	43.19370
33	635	60	43.19370
34	680	Missing	43.19370
35	681	Missing	43.19370
36	682	Missing	43.19370
37	683	Missing	43.19370
38	684	Missing	43.19370
39	685	Missing	43.19370
40	686	Missing	43.19370
41	687	48	43.19370
42	784	67	43.19370
43	857	47	43.19370
44	956	34	43.19370

Figure 3. Unweighted linear logit model and positive constrained WTP estimates with the complete case sample (n=936)

Binary Logit Model for Binary Choice
Dependent variable VOTE
Log likelihood function -615.34419
Restricted log likelihood -648.61267
Chi squared [9](P= .000) 66.53697
Significance level .00000
McFadden Pseudo R-squared .0512918
Estimation based on N = 936, K = 10
Inf.Cr.AIC = 1250.7 AIC/N = 1.336

VOTE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
WHOLE	.49764**	.24088	2.07	.0388	.02553	.96974
FIRST	.70411***	.18173	3.87	.0001	.34793	1.06029
SECOND	-.25725	.26477	-.97	.3312	-.77618	.26169
THIRD	.09561	.23749	.40	.6873	-.36986	.56108
FOURTH	.62564***	.23137	2.70	.0069	.17216	1.07913
AMOUNTW	-.00218*	.00117	-1.86	.0624	-.00448	.00011
AMOUNT1	-.00347***	.00096	-3.63	.0003	-.00534	-.00159
AMOUNT2	-.00425**	.00169	-2.52	.0117	-.00756	-.00095
AMOUNT3	-.00258**	.00127	-2.03	.0421	-.00507	-.00009
AMOUNT4	-.00311**	.00129	-2.41	.0158	-.00564	-.00059

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Jun 06, 2017 at 04:12:40 PM

WALD procedure. Estimates and standard errors for nonlinear functions and joint test of nonlinear restrictions.
VC matrix for the functions is singular.
Standard errors are reported, but the
Wald statistic cannot be computed.
Functions are computed at means of variables

WaldFcns	Function	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
WTPW	445.254**	192.7738	2.31	.0209	67.425	823.084
WTP1	319.045***	66.05767	4.83	.0000	189.574	448.515
WTP2	134.699***	38.47149	3.50	.0005	59.296	210.101
WTP3	287.397***	110.8228	2.59	.0095	70.188	504.606
WTP4	338.591***	109.5405	3.09	.0020	123.896	553.286
SUMPARTS	1079.73***	173.5641	6.22	.0000	739.55	1419.91

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Jun 06, 2017 at 04:12:41 PM

A Reply to John Whitehead's Replication Paper

by

William Desvousges[a*], Kristy Mathews[b], Kenneth Train[c]

[a]President, W.H. Desvousges & Associates, P.O. Box 99203, Raleigh, NC 27624

[b]Owner and Principal,104 McWaine Lane, Cary, NC 27513

[c]Adjunct 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.