Appalachian STATE UNIVERSITY

Department of Economics Working Paper

Number 15-07 | August 2015; Revised February 2016

Plausible Responsiveness to Scope in Contingent Valuation

John C. Whitehead
Appalachian State University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Plausible Responsiveness to Scope in Contingent Valuation

John C. Whitehead[1]

Department of Economics

Appalachian State University

Boone, North Carolina 28608

May 21, 2015 (Revised February 27, 2016)

[1] I thank Matt Interis and an anonymous journal referee for helpful comments.

Plausible Responsiveness to Scope in Contingent Valuation

Abstract. Plausible responsiveness to scope is a question of economic significance, in addition to statistical significance, of the scope test in contingent valuation. We briefly review the history of the scope test in order to place the current issue in the context of the BP Deepwater Horizon oil spill. As a result of the review we gain insights into how the issue of scope "adequacy" arose twenty years after it was first mentioned by the NOAA Panel on Contingent Valuation following the Exxon Valdez oil spill. We then provide a review of Desvousges, Mathews and Train (2012) who promote the adding-up test to identify inadequate responsiveness to scope adequacy. The adding-up test is a test of the construct validity of the contingent valuation method but is flawed as a measure of economic significance. We propose scope elasticity of willingness-to-pay as a measure of economic significance. A simulation suggests a likely range of elasticity estimates given linear and quadratic functional forms for the willingness-to-pay function. In order to illustrate the ease of implementation of scope elasticity within the context of the standard scope test we calculate scope elasticity with willingness-to-pay estimates from several studies, describe two studies that directly estimate scope elasticity and estimate scope elasticity with primary data from two studies. All of these empirical estimates of scope elasticity fall within the range of scope elasticity suggested by the simulation. Scope elasticity provides a practical way forward, relative to the adding up test, on the issue of economic significance of scope effects.

Key words: Contingent valuation method, scope test, elasticity, adding up test

1. Introduction

Stated preference surveys that use the contingent valuation method (CVM) elicit willingness-to-pay under various scenarios. One scenario is to vary the scope of environmental quality or natural resource allocation. The scope test is important in contingent valuation for two important reasons. Most practically, in theory but not often in practice, it can be used to estimate the total benefit curve. Smith (1984) makes the point that benefit-cost analysis in practice rarely identifies the policy level for net benefit maximization. Similarly, the CVM can be used to estimate benefits for a single policy level and these can be compared to costs. But implementation of the scope test can be used to determine the most efficient level of policy. Given a total cost curve, the total benefit (i.e., willingness-to-pay) curve could be used to estimate the optimal level of environmental quality or natural resource allocation instead of identifying levels that are more or less efficient than the status quo (e.g., Lindhjem et al. 2015).

The scope test is also an important construct validity test. Willingness-to-pay should be non-decreasing in the scope of environment quality or quantity of the natural resource allocation. Responsiveness to scope indicates that stated preferences conform to economic theory and lends validity to the CVM. On the other hand, failure of the scope test does not invalidate the CVM. The scope test is not a "crucial experiment" (Randall 1998) and failure could be due to diminishing marginal utility, substitution, behavioral anomalies, a poorly designed and/or executed survey or small sample sizes that reduce the power of the test. Any study should be assessed with a variety of validity and reliability tests (Whitehead and Haab 2013).

The NOAA Panel on Contingent Valuation (Arrow et al. 1993) stated that demonstration of "adequate" scope effects should be required for reliable measurement of willingness-to-pay for natural resource damage assessment. Desvousges, Mathews and Train (2012) consider the adding-up test to be an adequacy test. The purpose of this paper is to reframe the issue of scope sensitivity as one of economic significance and propose scope elasticity as an alternative to the adding up test. We first briefly review the history of the scope test in the context of the Exxon Valdez and BP Deepwater Horizon oil spills. We next consider the adding-up test and provide a critique of Desvousges, Mathews and Train (2012). We derive scope elasticity for two functional forms and conduct a Monte Carlo simulation to suggest a likely range. We provide a number of empirical examples to show the ease of implementation and understanding of scope elasticity.

2. A Brief History of the Scope Test

Mitchell and Carson (1989) first described the scope test as "part-whole bias." While scope tests had been performed prior to Mitchell and Carson, the controversy surrounding scope may have begun with the Kahneman (1986) and Kahneman and Knetsch (1992) "embedding" study. Kahneman and Knetsch (1992) find that willingness-to-pay is no different when a good is valued by itself and when it is valued as part of a larger bundle. Smith (1992) criticizes the survey and study design of Kahneman and Knetsch (1992) which fell short of best practice of the CVM.

Following the 1989 Exxon Valdez oil spill the state of Alaska commissioned a study to estimate the lost passive use values due to the spill with the intention to seek damages from Exxon. Carson et al. (1992, 2003) conduct a number of validity and reliability tests but do not conduct a scope test. In response, a team of researchers was assembled with funding from Exxon to critically evaluate the CVM and two volumes denouncing the contingent valuation method resulted (Hausman 1993, Desvousges et al. 2010). Among the papers that ultimately appeared in journals, Boyle et al. (1994) for waterfowl, and McFadden (1994) for wilderness areas, found that willingness-to-pay was insensitive to scope. There has been much criticism of the waterfowl and wilderness areas studies, including their inexpensive (at the time) survey modes (telephone and mall-intercept), open-ended valuation questions and small samples. Carson and Mitchell (1993) and Carson (1997) reexamine the Exxon-funded data and find that it does pass the scope test under certain conditions. Smith (1999) discusses survey design problems with the waterfowl survey.

The Exxon Valdez oil spill launched what has become known as the “CVM debate.” At the peak of the debate, the *Journal of Economic Perspectives* published a three paper symposium on the CVM. Portney (1994) introduced the issue and put it into context for economists who were unfamiliar with nonmarket valuation. Diamond and Hausman (1994), representing the “CVM critics,” review the anomalies and inconsistencies found in some past studies and conclude that, paraphrasing, no number is better than a CVM number. Internal scope tests, which were common in the literature, were described as weak by the CVM critics and split-sample external scope tests, also common but overlooked[2], were described as strong tests and suggested as necessary for reliable CVM surveys for natural resource damage assessments (Diamond and Hausman 1994). Hanemann (1994) argues against many of the points made by Diamond and Hausman and critiques Kahnemann and Knetsch (1992) and Desvousges et al. (2010). Carson and Mitchell (1996) clarified the terms embedding, part-whole bias and internal and external scope tests. Many split-sample external tests of the CVM were subsequently conducted (e.g., Whitehead, Haab and Huang, 1998; Berrens et al. 2000).

[2] See Carson (1997) on this point.

Following the BP Deepwater Horizon oil spill, the *Journal of Environmental Perspectives* published a second symposium on the CVM. Kling, Zhao and Phaneuf (2012) review the literature and conclude that CVM studies, especially those conducted since the beginning of the CVM debate, tend to pass the scope test. Carson (2012) reviews evidence that CVM surveys tend to pass the scope test and survey design issues that can lead to insensitivity to scope. Hausman (2012) argues that since a few selected studies do not pass the scope test and few studies have conducted an adding-up test, among other issues, then the CVM is "hopeless." Haab et al. (2013) comment on Hausman (2012) by broadening the literature review and arguing that most CVM studies, summarized in several meta-analyses, tend to pass a scope test.[3]

3. Plausibility and the Adding Up Test

The NOAA Panel on Contingent Valuation (Arrow et al. 1993) was established in 1991 to arbitrate between the views of the CVM proponents and critics and develop a set of guidelines for the conduct of the CVM for natural resource damage assessment. The NOAA Panel described the scope test as a test of rationality: "Usually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual. Also, if marginal or incremental willingness to pay for additional amounts does decline with the amount already available, it is usually not reasonable to assume that it declines very abruptly." The NOAA Panel observed that some CVM studies fail to pass this test: "Different but similar samples of respondents are asked about their willingness to

[3] See Desvousges, Mathews and Train (2016) and Haab et al. (2016) for a comment and reply.

pay for prevention of environmental damage scenarios that are identical except for their scale: different numbers of seabirds saved, different numbers of forest tracts preserved from logging, etc. It is reported that average willingness to pay is often substantial for the smallest scenario presented but is then substantially independent of the size of the damage averted, rising slightly if at all for large changes in size." In terms of guidance for future surveys they included "inadequate responsiveness to the scope of the environmental insult" as one item in a list of "maladies" that would render a CVM survey unreliable. The burden of proof fell to the researcher.

The NOAA Panel left "inadequate" open to interpretation. In a memo to the U.S. EPA, published in a report critical of NOAA guidelines for the CVM, a subset of the NOAA panel, Arrow, Leamer, Schuman and Solow (Arrow et al. 1994) attempted to clarify adequacy when it had been confused with statistical significance: "Had the panel thought that something as straightforward as statistical measurability were the proper way to define sensitivity, then we would (or should) have opted for language to that effect. A better word than 'adequate' would have been 'plausible': A survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult. This, of course, is a judgment call, and cannot be tested in a context-free manner …." Arrow et al. (1994) left "plausible" open to interpretation. Synonyms for adequate include sufficient while synonyms for plausible includes believable. Sufficient responsiveness to scope suggests a threshold that must be met by the data. A believable scope effect is less restrictive, suggesting that the magnitude should be within the realm of possibility. With either, consideration of scope adequacy or plausibility is similar to the call to consider economic significance in addition to statistical significance in all fields of economics (McCloskey and Ziliak 1996). While economists in other fields who enjoy large data sets may be guilty of ignoring economic significance, CVM researchers routinely report the magnitude of economic effects by presenting willingness-to-pay estimates that can be judged on their plausibility or believability.

In the next section we discuss how Desvousges, Mathews and Train (2012) consider the adding up test as a test of adequacy. Diamond (1996) provides an example in his footnote 14: "As examples of possible adding-up tests, consider variations on two recent surveys. Schulze et al. [11] used two surveys to ask for WTP for partial and complete cleanups of the Upper Clark Fork River Basin in Montana. For an adding-up test, a third survey would describe a partial cleanup and describe the government as already committed to it, with the costs to be borne as described in the existing survey. The survey would then describe a complete cleanup and ask for WTP to enhance the cleanup from partial to complete. The mean WTP response from this question plus the mean WTP for partial cleanup should be almost exactly the same as the mean WTP for complete cleanup. One could test for the statistical significance of any difference that was found." Diamond (1996) formalized the adding-up test in the context of the waterfowl and wilderness areas studies but does not provide an empirical test.[4] To understand the adding-up test, consider goods A, B and C where $A = B + C$. A valuation study passes the adding up test if $WTP_A = WTP_B + WTP_C$. Due to substitution and income effects, the adding up test must be implemented by separately eliciting WTP_A, WTP_B and $WTP_C|B, Y - T_B$ where B indicates the amount of B purchased, Y is income and T_B is the cost of B. The empirical test is for whether

[4] See Diamond et al. (1993) for an empirical test using the wilderness area data. It is not clear why these results do not appear in Diamond (1996) but consider footnote 29 in Hanemann (1994).

$WTP_A = WTP_B + WTP_C|B, Y - T_B$. In contrast, a scope test is implemented by separately eliciting WTP_A and WTP_B and testing whether $WTP_A > WTP_B$.

Base (e.g., partial cleanup) and scope (e.g., complete cleanup) scenarios are generally straightforward to describe to survey respondents. As described by Diamond (1996) developing the adding-up scenario complicates survey development. First, it is an ex-post counterfactual scenario. Respondents must be convinced that a currently nonexistent government program has been funded and implemented and that their budget has been reduced by the cost. Second, the adding-up test raises survey costs. In order to develop a convincing counterfactual, additional focus groups and pretesting are necessary. Failure to pursue these additional survey development tasks in order to provide a convincing counterfactual scenario could lead to failure of the adding-up test. In addition to increased survey development costs, the adding-up test increases survey costs by adding another treatment that increases sample size. Beyond the monetary costs, the adding-up test treatment imposes an opportunity cost. The additional monetary costs could be employed in an alternative scope treatment to elicit another point on the total benefit curve, helping to identify the level of the policy where net benefits are maximized.

One adding-up test appeared in the journal literature between the time of the Exxon Valdez and BP Deepwater Horizon oil spills. Bateman et al. (1997) elicit independent valuations of vouchers for a restaurant meal and the two components of the meal (the main course and coffee plus dessert) in a laboratory experiment. They follow an adding-up test experimental design and find that the value of the two components exceed the value of the whole, indicating part-whole bias exists for well-known market goods. They conclude: "We have found clear evidence of [part-whole bias] for private consumption goods, bought and sold using incentive compatible mechanisms. Our results suggest, contrary to most previous interpretations of [part-whole bias], that this phenomenon may not be attributable simply to problems with the CV method, or with specific applications of that method. Instead, it may be a symptom of some fundamental property of individual's preferences which conventional consumer theory does not allow for." In a controlled laboratory environment the adding-up test did not pass for well-known market goods and the authors interpret the result as a behavioral anomaly (Bateman et al. 1997).[5]

4. Desvousges, Mathews and Train (2012)

Desvousges, Mathews and Train (2012) review the scope effects literature and classify studies based on whether they pass or fail the scope test based on statistical significance. If all of the differences in scope published in the study are statistically significant then the verdict is that the study passes the scope test. If some of the tests pass, but not all, then the evidence is considered mixed. If none of the scope tests pass then the study is classified as a fail. Of 109 studies, 36% pass the scope test, 15% fail the test and 49% have mixed evidence. The classification system potentially obscures a number of fine points. For example, Rollins and Lyke (1998) explicitly test for diminishing marginal utility as the scope of the good increases. They find sensitivity to scope at lower levels of provision of the good and lack of sensitivity at higher levels of provision, as theory suggests due to diminishing marginal value. Giraud, Loomis and Johnson (1999) find that their study passes a "strong" external (i.e., split-sample) scope test but not a "weak" internal (i.e., paired comparison) test. Whitehead and Cherry (2007) investigate different

[5] See also Elbakidze and Nayga (2015) and Desvousges, Mathews and Train (2015) for similar negative results.

empirical models with alternative hypothetical bias correction methods and find that some specifications exhibit sensitivity to scope while others do not. In each of these examples, it is more appropriate to classify the results as pass, since a preferred specification passes the scope test and sensitivity analysis is conducted to determine when the data does not pass the test. While a full exploration is beyond the scope of this paper, other studies where the scope test passed a statistical significance test in a preferred specification may have been uncharitably labeled as "mixed" by Desvousges, Mathews and Train (2012). Their classification system results in an overly pessimistic conclusion about the statistical significance of the scope test.

Next, Desvousges, Mathews and Train state that they: "… examined all of the studies … to determine what, if anything, they reveal about adequacy of response to scope. Amazingly, there is little information to be obtained. In most studies, the original and reduced environmental goods that are specified for the scope test differ in ways that prevent an assessment of the magnitude of response." They continue: "For example, Berrens et al. (2000) find that the average value of saving one fish species in one river is $57 while the value of saving 11 fish species in four rivers is $74, which might appear to be inadequate response (i.e., "look fishy") but might also be the result of highly diminishing marginal utility of saving fish." In contrast to Desvousges, Mathews and Train's (2012) use of the example, it shows how CVM studies typically provide the information necessary to assess the economic significance of the scope effect as we show below. Desvousges, Mathews and Train (2012) do not pursue a more systematic and detailed review to assess the economic significance of the scope effect in previous studies.

Desvousges, Mathews and Train (2012) then assert that the adding-up test is an adequacy test: "… the hypothesis is tested of whether the sum of the first two WTPs is equal to the third. The issue of adequacy is addressed by this procedure, since the response to scope is inadequate if the sum of the parts exceeds the whole." Of the 109 scope studies, only three provide enough information to meet the Desvousges, Mathews and Train (2012) requirements for testing scope adequacy. In two of these studies adding-up is rejected. The third study is the Chapman et al. (2009) natural resource damage assessment which conducted a CVM study with only base and scope scenarios. In contrast to Diamond (1996) who describes three different willingness-to-pay scenarios and explicit discussion of payment (T_B in the previous section) needed for an adding up test, Desvousges, Mathews and Train (2012) interpret Chapman et al. (2009) as containing a base willingness-to-pay with incremental parts. Willingness-to-pay in the adding up scenario ($WTP_C|B, Y - T_B$ in the previous section) is inferred from the residual of the base and scope willingness-to-pay estimates and Train (2012) assesses income effects statistically to suggest that T_B is not a factor. They argue that Chapman et al. (2009) passes the scope test but does not pass their interpretation of the adding up test.

Desvousges, Mathews and Train (2012) conclude: "One thing seems clear to us. The NOAA Panel was concerned with the possibility of 'inadequate responsiveness to the scope of the environmental insult.' The standard scope test does not address this concern, since it tests for statistical significance rather than adequacy of magnitude. In nearly all past applications of the test, the design of the study does not permit an evaluation of adequacy. And in the one study that permits such an evaluation while passing the standard scope test, the response to scope is demonstrably inadequate – which shows that the standard scope test does not provide a reliable indicator of adequacy." Desvousges, Mathews and Train (2012) consider the standard scope test a statistical significance test and equate failure of the adding up test with scope inadequacy. This ignores information available in many studies to assess the economic significance of the scope test.

5. Scope Elasticity of Willingness-to-pay

Elasticity is often used to assess the plausibility of the change in one economic variable in response to the change in another. For example, an own-price elasticity of demand equal to zero indicates a vertical demand curve and violation of the principle of demand. Any positive own-price elasticity (in absolute value) is a plausible demand curve. Economists routinely refer to "inelastic" demand and attempt to find the economic factors responsible for the limited response (e.g., few substitutes, small budget share, short run time period). Similarly, given a willingness-to-pay function, $WTP(Q)$, scope point elasticity of willingness-to-pay could be measured as $\varepsilon_Q = \frac{dWTP}{dQ}\frac{Q}{WTP}$, where Q is a quantitative measure of scope change.[6] Many CVM studies include only base and scope scenarios, which is not enough split sample treatments to estimate a continuous willingness-to-pay function and the point elasticity. In this case the arc elasticity could be used: $\varepsilon_Q = \frac{\Delta WTP}{\Delta Q}\frac{\bar{Q}}{\overline{WTP}}$; where the horizontal bar represents the midpoint of a line segment.

---

[6] Amiran and Hagen (2010) develop a formal model of scope with bounded utility functions and show that relatively small scope effects are not inconsistent with economic theory. The size of the scope effect in their model depends explicitly on the substitutability between market and nonmarket goods.

With typical willingness-to-pay functions it can be expected that scope elasticity will range from zero to one. Consider the linear functional form with constant marginal value, $WTP = a + bQ$, where $a \geq 0$, $b \geq 0$ and scope elasticity is $0 \leq \frac{bQ}{a+bQ} \leq 1$. The elasticity is equal to one if $a = 0$ and is less than one if $a > 0$.[7] Studies that find a statistically insignificant scope effect, $b = 0$, will have scope elasticity equal to zero. The quadratic functional form allows diminishing marginal value: $WTP = c + dQ + eQ^2$, where $c \geq 0$, $d \geq 0$, $e \leq 0$ and scope elasticity is $0 \leq \frac{dQ+2eQ^2}{c+dQ+eQ^2} < 1$.[8] With statistically significant scope effects and diminishing marginal value, $d > 0, e < 0$, scope elasticity is always less than one since $2e < c + e$. In order to better understand the likely range of scope elasticity with these two functional forms we conduct a Monte Carlo simulation where a, b, c, d and Q are randomly drawn from a uniform distribution that ranges from 1 to 100 and e is randomly drawn from a uniform distribution that ranges from -0.00005 to -0.005. The mean WTP is 2637 and 2889 with the linear and quadratic functional forms over 1000 random draws. The mean scope elasticity is 0.927 with a 95% confidence interval of 0.630 and 0.998 with the linear functional form. The mean scope elasticity with the quadratic functional form is 0.708 with a 95% confidence interval of 0.177 and 0.971. This simulation is

[7] Carson (1997) estimates a willingness-to-pay model with the 10% trimmed wilderness area data from Diamond et al. (1993) and finds that scope elasticity, as defined here, is equal to one.

[8] In footnote 5 Diamond (1996) suggests that marginal willingness-to-pay should be increasing in scope suggesting an elasticity greater than one. Hanemann (2004) addresses this by critiquing Diamond's assumptions.

not conclusive but it does suggest the range of plausible elasticities that might be expected from willingness-to-pay functions with statistically significant scope effects.

Considering the Berrens et al. (2000) "fishy" example from Desvousges, Mathews and Train (2012) mentioned above the scope arc elasticity is 0.16 if the number of protected fish species is the relevant quantity.[9] If the number of rivers with protected in-stream flows is the relevant quantity, the scope arc elasticity is $0.22 = \left(\frac{74-57}{4-1}\right)\left(\frac{(4+1)/2}{(74+57)/2}\right)$. Desvousges, Mathews and Train (2012) employ the Turnbull estimator which generates the smallest difference in willingness-to-pay from Berrens et al. (2000). The median willingness-to-pay from three parametric models is "approximately" \$25 and \$55 for one fish (one river) and eleven fish (four rivers). If these willingness-to-pay estimates are used then the scope arc elasticities are 0.45 and 0.63 when fish and rivers are the relevant quantities.

The plausibility of the scope effect can be assessed for many studies in the literature. For example, consider three recently published studies. Lindhjem et al. (2015) finds statistically significant differences in willingness-to-pay for an increase in preserved forests from a 1.4% baseline to 2.8% and from 1.4% to 10%. Willingness to pay for the increase to 2.8% is \$693 and willingness-to-pay for the increase to 10% is \$871. The scope arc elasticity is 0.20. Martin-Ortega, Mesa-Jurado and Berbel (2015), examining ordering effects, find statistically significant differences in willingness-to-pay for an increase in water reliability from 50% to 90%. Willingness-to-pay for a 50% increase in two ordering scenarios is \$46 and \$50 and willingness-to-pay for a 90% increase is \$77 and \$72. The scope arc elasticities are 0.88 and 0.63. Longo,

[9]The calculation is $0.16 = \left(\frac{74-57}{11-1}\right)\left(\frac{(11+1)/2}{(74+57)/2}\right)$.

Hoyos, and Markandya (2015), also examining ordering effects, estimate the value of reducing greenhouse gas emissions from three policies that have different effectiveness. The Turnbull willingness-to-pay estimates are $62, $99 and $152 for reductions of 0.5%, 4% and 16% in one sequence of valuation questions and $49, $89 and $109 in another sequence. Scope elasticity of willingness-to-pay ranges from 0.15 to 0.35.

Two studies provide estimates of scope elasticity in their willingness-to-pay models. Richardson and Loomis (2009) consider the scope effect with secondary CVM data in a meta-analysis of 67 willingness-to-pay estimates of threatened and endangered species protection. They find that the magnitude of the scope effect is a statistically significant determinant of willingness-to-pay. The linear "full model for benefit transfer" (Table 5) scope elasticity is 0.27 with a 95% confidence interval of 0.03 to 0.51. The scope elasticity from the double log "full model for benefit transfer," which has a better statistical fit, is 0.71 with a 95% confidence interval of 0.38 and 1.04. Metcalfe et al. (2012) estimate the value of water quality improvements in England and Wales with primary data. The scope treatments are estimated with the existing water quality as the base and two randomly assigned improvements. The scope elasticity is 0.68 with a 95% confidence interval of 0.14 to 1.22. In a model with covariates the scope elasticity is 0.51 with a 95% confidence interval of -0.09 to 1.11. The discussion in Metcalfe et al. (2012) implies that this result is plausible.

In order to estimate scope elasticity with CVM data, we consider two data sets from the scope test literature reviewed by Desvousges, Mathews and Train (2012). First, Banzhaf et al. (2006) estimate the benefits of improved water quality in Adirondack State Park to New York state residents. We construct the data from Table 4 found in a preliminary version of the paper

(Banzhaf et al. 2004). We use the data from the four bid levels with large sample sizes: \$25, \$90, \$150 and \$250.[10] The level of environmental quality for the base and scope scenarios are 20% and 40% increase in "lakes that support fish in ten years." The estimated model is $WTP = 135 + 90Q$, where Q is equal to one for a 40% increase and zero otherwise.[11] The arc elasticity is 0.76 with a 95% confidence interval of 0.42 and 1.10. Whitehead and Cherry (2007) elicited referendum votes to four bid amounts (\$5, \$15, \$30 and \$50) and three levels of air quality (health and visibility) improvement (2%, 10% and 20%). The estimated double log model is $lnWTP = 1.18 + 0.56lnQ$.[12] The scope elasticity is the coefficient on the scope variable, 0.56, with a 95% confidence interval of .055 and 1.06.

[10] We do not use 50 base case observations at three bid values (\$35, \$85, \$200). There are no observations with the scope scenario with these bid values.

[11] Willingness-to-pay and scope elasticity are estimated with the censored probit model and the standard errors are estimated with the Delta Method (Cameron and James 1987, Cameron 1991). Each of the censored probit coefficient estimates are statistically significant. These results are available upon request.

[12] Willingness-to-pay and scope elasticity are estimated with the censored probit model and the standard errors are estimated with the Delta Method (Cameron and James 1987, Cameron 1991). Each of the coefficient estimates are statistically significant. The double log model generates the best statistical fit. The scope elasticity from linear and quadratic models are not statistically different from one. These results are available upon request.

6. Conclusions

The adding-up test is a useful construct validity test but very few researchers have chosen to employ it. The reasons may include that it has higher costs and lower benefits relative to a basic scope test. The adding-up test is less useful for determining if scope effects are plausible. A test for plausibility is a test for economic significance that should be assessed with changes in the magnitude of willingness-to-pay relative to changes in the magnitude of the scope variable. We propose scope elasticity of willingness-to-pay as a measure of the economic significance of scope that can be compared across studies to assess plausibility. Scope elasticity is a straightforward measure of economic significance and many existing studies will contain the necessary information. We examine scope elasticity from a number of existing studies and find that in each study scope elasticity is within a range suggested by a simulation. Where confidence intervals can be developed, scope elasticity is statistically different from zero but not statistically different from one. While our sample of studies is purposively small, all of the estimates exhibit plausible responsiveness to scope.

CVM studies that pass the scope test produce results that are most useful for policy analysis. CVM studies that do not pass the scope test should be critically examined for behavioral anomalies (Heberlein et al. 2005) and other issues before the CVM is determined to be a valuation method that cannot measure preferences. One step in this direction is to conduct a more thorough review of scope effect literature. Similar to Richardson and Loomis (2009), a meta-analysis should be performed with a secondary data set that includes as the dependent variable the magnitude of willingness-to-pay for each of the scope levels considered in each study. The other independent variables should include study characteristics such as valuation question format, survey mode, quantitative or qualitative descriptions of scope, use of visual aids, geographic, temporal or other variant of scope and other standard control variables. Relative to Desvousges, Mathews and Train's (2012) review that focuses on statistical significance and the adding up test, this is the sort of study needed to fully assess the conditions under which the magnitude of scope effects are plausible.

References

Amiran, Edoh Y., and Daniel A. Hagen. "The scope trials: Variation in sensitivity to scope and WTP with directionally bounded utility functions." Journal of Environmental Economics and Management 59, no. 3 (2010): 293-301.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. “Report of the NOAA Panel on Contingent Valuation” (1993).

Arrow, Kenneth, Edward E. Leamer, Howard Schuman and Robert Solow. Appendix D in “Comments on Proposed NOAA/DOI Regulations on Natural Resource Damage Assessment,” U.S. Environmental Protection Agency (October 1994).

Banzhaf, Spencer, Dallas Burtraw, David Evans, and Alan Krupnick. "Valuation of Natural Resource Improvements in the Adirondacks." Resources for the Future (September 2004).

Banzhaf, H. Spencer, Dallas Burtraw, David Evans, and Alan Krupnick. "Valuation of natural resource improvements in the Adirondacks." Land Economics 82, no. 3 (2006): 445-464.

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. "Does part–whole bias exist? An experimental investigation." The Economic Journal 107, no. 441 (1997): 322-332.

Berrens, Robert P., Alok K. Bohara, Carol L. Silva, D. Brookshire, and Michael Mckee. "Contingent values for New Mexico instream flows: With tests of scope, group-size reminder and temporal reliability." Journal of Environmental Management 58, no. 1 (2000): 73-90.

Boyle, Kevin J., William H. Desvousges, F. Reed Johnson, Richard W. Dunford, and Sara P. Hudson. "An investigation of part-whole biases in contingent-valuation studies." Journal of Environmental Economics and Management 27, no. 1 (1994): 64-83.

Carson, Richard T. “Contingent valuation surveys and tests of insensitivity to scope.” Chapter 6 (Pages 101-125) in Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods, Editors: R. J. Kopp, W. W. Pommerehne, N. Schwarz Springer Netherlands, 1997.

Carson, Richard T., and Robert Cameron Mitchell. "The issue of scope in contingent valuation studies." American Journal of Agricultural Economics 75, no. 5 (1993): 1263-1267.

Carson, Richard T., and Robert Cameron Mitchell. “Sequencing and Nesting in Contingent Valuation Surveys.” Journal of Environmental Economics and Management 28, no. 2 (1995): 155-173.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. No. 6984. University Library of Munich, Germany, 1992.

Carson, Richard T., Robert C. Mitchell, Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. “Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill.” Environmental and Resource Economics 25, no. 3 (2003): 257-286.

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Boyle, K. J., Hudson, S. P., Wilson, K. N. (2010). Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy, 2nd ed. RTI Press publication No. BK-0001-1009. Research Triangle Park, NC: RTI International.

Desvousges, William, Kristy Mathews, and Kenneth Train. "Adequate responsiveness to scope in contingent valuation." Ecological Economics 84 (December 2012): 121-128.

Desvousges, William, Kristy Mathews, and Kenneth Train. "An Adding-up Test on Contingent Valuations of River and Lake Quality." Land Economics 91 (August 2015): 556-571.

Desvousges, William, Kristy Mathews, and Kenneth Train. "From Curious to Pragmatically Curious: Comment on "From Hopeless to Curious? Thoughts on Hausman's 'Dubious to Hopeless' Critique of Contingent Valuation"." Applied Economic Perspectives and Policy 38, no. 1 (2016): 174-182.

Diamond, Peter A. 1996. "Testing the Internal Consistency of Contingent Valuation Surveys." Journal of Environmental Economics and Management 30: 337–347.

Diamond, Peter A., and Jerry A. Hausman. "Contingent valuation: Is some number better than no number?" Journal of Economic Perspectives 8, no. 4 (1994): 45-64.

Diamond, Peter A., Jerry A. Hausman, Gregory K. Leonard, and Mike A. Denning. 1993. "Does Contingent Valuation Measure Preferences? Experimental Evidence." In Contingent Valuation, A Critical Assessment, ed. J.A. Hausman, 41–89. Amsterdam: Elsevier.

Elbakidze, L., and R. Nayga. 2015. Validating Consistency of Non-hypothetical Experimental Auction Data: Application of the Adding-up Test in a Multi-unit Setting. Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's Annual Meeting, San Francisco, CA.

Giraud, K. L., John B. Loomis, and R. L. Johnson. "Internal and external scope in willingness-to-pay estimates for threatened and endangered wildlife." Journal of Environmental Management 56, no. 3 (1999): 221-229.

Haab, Timothy C., Matthew G. Interis, Daniel R. Petrolia, and John C. Whitehead. "From hopeless to curious? Thoughts on Hausman's 'dubious to hopeless' critique of contingent valuation." Applied Economic Perspectives and Policy 35, no. 4 (2013): 593-612.

Haab, Timothy C., Matthew G. Interis, Daniel R. Petrolia, and John C. Whitehead. "Interesting Questions Worthy of Further Study: Our Reply to Desvousges, Mathews, and Train's (2015) Comment on Our Thoughts (2013) on Hausman's (2012) Update of Diamond and Hausman's (1994) Critique of Contingent Valuation." Applied Economic Perspectives and Policy 38, no. 1 (2016): 183-189.

Hausman, Jerry. "Contingent valuation: from dubious to hopeless." The Journal of Economic Perspectives 26, no. 4 (2012): 43-56.

Heberlein, Thomas A., Matthew A. Wilson, Richard C. Bishop, and Nora Cate Schaeffer. "Rethinking the scope test as a criterion for validity in contingent valuation." Journal of Environmental Economics and Management 50, no. 1 (2005): 1-22.

Kahneman, Daniel. "Comments." In Valuing Environmental Goods: An Assessment of the Contingent Valuation Method, edited by Ronald G. Cummings, David S. Brookshire and William D. Schulze. Totowa, NJ: Rowman and Allanheld (1986).

Kahneman, Daniel, and Jack L. Knetsch. "Valuing public goods: the purchase of moral satisfaction." Journal of Environmental Economics and Management 22, no. 1 (1992): 57-70.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. "From Exxon to BP: Has some number become better than no number?" Journal of Economic Perspectives 26, no. 4 (2012): 3-26.

Lindhjem, Henrik, Kristine Grimsrud, Ståle Navrud, and Stein Olav Kolle. "The social benefits and costs of preserving forest biodiversity and ecosystem services." Journal of Environmental Economics and Policy 4, no. 2 (2015): 202-222.

Longo, Alberto, David Hoyos, and Anil Markandya. "Sequence Effects in the Valuation of Multiple Environmental Programs Using the Contingent Valuation Method." Land Economics 91, no. 1 (2015): 20-35.

Martin-Ortega, Julia, M. Azahara Mesa-Jurado, and Julio Berbel. "Revisiting the Impact of Order Effects on Sensitivity to Scope: A Contingent Valuation of a Common-Pool Resource." Journal of Agricultural Economics 66, no. 3 (2015): 705-726.

McCloskey, Deirdre N., and Stephen T. Ziliak. "The standard error of regressions." Journal of Economic Literature 34, no. 1 (1996): 97-114.

McFadden, Daniel. "Contingent valuation and social choice." American Journal of Agricultural Economics 76, no. 4 (1994): 689-708.

Metcalfe, Paul J., William Baker, Kevin Andrews, Giles Atkinson, Ian J. Bateman, Sarah Butler, Richard T. Carson, Jo East, Yves Gueron, Rob Sheldon and Kenneth Train. "An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales." Water Resources Research 48, no 3 (2012).

Randall, Alan. "Beyond the crucial experiment: mapping the performance characteristics of contingent valuation." Resource and Energy Economics 20, no. 2 (1998): 197-206.

Richardson, Leslie, and John Loomis. 2009. “The Total Economic Value of Threatened, Endangered and Rare Species: An Updated Meta-Analysis.” Ecological Economics 68 (5): 1535-1548.

Rollins, Kimberly, and Audrey Lyke. "The case for diminishing marginal existence values." Journal of Environmental Economics and Management 36, no. 3 (1998): 324-344.

Smith, V. Kerry. "Environmental policy under Reagan's executive order: the role of benefit-cost analysis." (1984).

Smith, V. Kerry. "Arbitrary values, good causes, and premature verdicts." Journal of Environmental Economics and Management 22, no. 1 (1992): 71-89.

Smith, V. Kerry. "Of birds and books: more on hypothetical referenda." Journal of Political Economy 107, no. 1 (1999): 197-200.

Train, Kenneth. “Comments on Chapman et al. (2009): Inadequate Response to Scope while Passing the Scope Test,” January 24, 2012.

Whitehead, John C., and Todd L. Cherry. "Willingness-to-pay for a green energy program: a comparison of ex-ante and ex-post hypothetical bias mitigation approaches." Resource and Energy Economics 29, no. 4 (2007): 247-261.

Whitehead, John C., and Timothy C. Haab. "Contingent Valuation Method," in Encyclopedia of Energy, Natural Resource, and Environmental Economics, ed. Jason Shogren, Elsevier, 2013.

Whitehead, John C., Timothy C. Haab, and Ju-Chin Huang. "Part-whole bias in contingent valuation: Will scope effects be detected with inexpensive survey methods?." Southern Economic Journal (1998): 160-168.