Sensitivity to Scope: Evidence from a CVM Study of Wetlands

Daniel J. Mullarkey
Economic Research Service, USDA
1800 M Street NW, Room 4087
Washington, DC 20036
Phone: 202-649-5599
Fax: 202-649-5756
Email: danielm@econ.ag.gov

Richard C. Bishop, Chair
Dept. of Agricultural and Applied Economics
University of Wisconsin-Madison
427 Lorch Street
Madison, WI 53706
Phone: 608-262-8966
Fax: 608-262-4376
Email: bishop@aae.wisc.edu

Presented at:

American Agricultural Economics Association
Annual Meeting
Nashville, TN, August 8-11, 1999

Presenter and primary contact: Daniel J. Mullarkey[1]

Abstract: Wetlands valuation is a situation in which CVM studies might be expected to fail scope tests. This paper reports results from a split-sample CVM study of Wisconsin wetlands. The survey employed a multiple-bounded, polychotomous-choice format, and compared WTP distributions using the method of convolutions. The survey demonstrated sensitivity to scope.

---

[1] Daniel J. Mullarkey is an Economist at the Economic Research Service, USDA, and Richard C. Bishop is chair of the Department of Agricultural and Applied Economics at the University of Wisconsin-Madison. The views expressed in this article do not necessarily represent the policies or views of the U.S. Department of Agriculture. Copyright 1999 by Daniel J. Mullarkey and Richard C. Bishop. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Introduction

The debate over whether the Contingent Valuation Method (CVM) can provide valid estimates of economic value focuses in part on whether willingness to pay (WTP) estimates are sensitive to the scope of the good being valued. According to economic theory, if good A contains all of the attributes of good B (in at least the same levels), and also includes positive levels of at least one additional attribute, then good A should be valued higher than good B. This allows a testable hypothesis for CVM applications, which can be used to provide evidence on the theoretical construct validity of the survey in question (Bishop, Champ and Mullarkey, 1995). The issue of scope insensitivity has been championed by critics of CV (Kahneman and Knetsch, 1992; Diamond et al., 1993; Desvousges et al., 1993; and Schkade and Payne, 1994). These studies have been criticized elsewhere (Smith, 1992; Hanemann, 1994; Carson, 1995; Mullarkey, 1997) as lacking content validity, with primary objections being that the goods to be valued are poorly defined, and that survey design features and econometric techniques have not been state-of-the-art. In reviewing past studies, Carson (1995) finds considerable evidence of sensitivity to scope, with the exceptions being the studies cited above. Despite the flaws in these surveys, sensitivity to scope is a legitimate issue to raise, and the increased attention will result in better CV studies.

Wetlands valuation represents a perfect situation in which to explore this issue. Wetlands valuation has historically been piecemeal and incomplete (Mullarkey, 1997), and there is interest in determining whether CVM can provide better information about total values for the services of particular wetlands.[2] However, values for many wetlands will contain components of nonuse

[2] Good information on values for wetlands remains a priority despite recent reductions in the rate of loss of wetlands. One of the major tools in this push toward "no net loss" of wetlands

and indirect use values; the diverse services provided by wetlands are relatively unfamiliar to many people; and there is considerable scientific uncertainty surrounding the ecological functions performed by wetlands. These are factors that have been identified as being troublesome for CVM, leading to the suggestion that survey participants will express their values for wetlands in general rather than the wetlands of interest to a particular survey. This paper reports on a CVM survey designed to test sensitivity to scope, with the application being a wetlands preservation project.

Case Study

The case study chosen for this research was a highway expansion project in northwest Wisconsin that required the filling of 110 acres of wetlands. The Wisconsin Department of Transportation (WDOT) had decided to expand a 44-mile stretch of U.S. Highway 53 (USH 53). Plans called for the two new lanes to traverse 47 acres of tributary system wetlands and 63 acres of isolated basins. The USH 53 wetlands were located in an area that has an abundance of wetlands, and were not known to contain any threatened or endangered species. These are classic examples of the type of unspectacular wetlands that continue to be lost in small but steady increments, and about which we have very little valuation information.

These losses are being mitigated by the creation of 220 acres of isolated basins at different sites along the highway. In accordance with the cooperative agreement between the WDOT and the Wisconsin Department of Natural Resources, a 2-1 replacement ratio was chosen to reflect the somewhat imperfect success of wetland creation attempts to replace all of the lost

---

is mitigation, where wetlands are restored or created in one location to offset loss of wetlands in a nearby location. In such cases, where the goal is a fair trade of wetland services, it is important

wetland services. This project was selected in large part because the Environmental Impact Statement prepared by the WDOT contained the most detailed ecological data set on the services provided by any wetlands in Wisconsin that faced the current threat of development.

A representative subset of the 130 wetlands was evaluated on a service-by-service basis using the Wetland Evaluation Technique (WET). Sixteen wetlands were rated for eleven services: groundwater recharge, groundwater discharge, floodflow alteration, sediment stabilization, sediment and toxicant retention, nutrient removal and transformation, primary production export, aquatic diversity and abundance, wildlife breeding habitat, wildlife migration habitat, and wildlife wintering habitat. WET returns a qualitative rating of high, medium, low or uncertain for each service evaluated. A rating of "high" for a particular function implies that there is a high probability that the wetland provides that service, rather than implying that the wetland provides a large quantity of that service.

Reflecting the diversity amongst even wetlands in the same geographic area, the WET ratings for the USH 53 wetlands varied quite a bit. There were clear differences between the isolated basins and the tributary system wetlands, which necessitated describing the two types separately. There were also differences within each type, which complicated the task of providing a general description of the quality of each service. Providing useful information for eleven services and two types of wetlands was the central challenge to developing a survey with strong content validity.[3] After three focus groups, a pretest, and consultations with WDOT wetland specialists, we condensed the eleven services to six. In response to feedback from focus group participants, we used one table with columns for each type of wetland, rather than having

---

to understand both the value of the original wetlands and the value of the mitigation wetlands.

separate table for each type of wetland. The majority of focus group participants wanted quantitative information, thus we provided percentages of each type of wetland rating high, moderate, or low for each service. The final information set of reference levels for the wetland services in provided in Figure 1.

Figure 1: Table of Wetland Services

BENEFIT	RIVER SYSTEM	ISOLATED BASINS
Floodflow Alteration	These wetlands are good at slowing flood flows. This reduces erosion and damage to downstream areas.	Isolated basins are also good at slowing floods, absorbing water, and reducing the potential for flood damage.
Water Purification	Some river system wetlands purify water by filtering out pollutants (sediment, nutrients, and some toxic substances).	The isolated basins effectively protect surface and groundwater supplies by removing pollutants (sediment, nutrients, and some toxic substances).
Fish and Other Aquatic Animals	60% of the river system wetlands can support fish and aquatic species, but the other 40% cannot.	The isolated basins generally are <u>not</u> suitable habitat for fish and other aquatic populations.
Wildlife Breeding Habitat	These river system wetlands are good breeding habitat for a variety of wildlife.	For breeding habitat, 45% of the isolated basins rate low, 22% medium, and 33% high.
Bird Migration Habitat	These wetlands provide good habitat for migrating birds.	For migration habitat, 56% of the isolated basins rate low, 22% medium and 22% high.
Bird Wintering Habitat	Most of these river system wetlands are good wintering habitat; the 30% that freeze over are not.	These isolated basins freeze in the winter, which makes them poor habitat for wintering birds.

[3] Developing a survey that was perceived as neutral between the choice of development or preservation was also challenging, and the focus groups were quite useful in this endeavor.

One decision that needs to be made in every CVM survey is which value elicitation device, or question format, to employ. Different types of valuation questions can produce different welfare estimates (Welsh and Poe, 1998). The question format selected for this survey was the multiple-bounded polychotomous-choice (MB-PC) framework (see Welsh and Bishop, 1993, and Welsh and Poe, 1998). Multiple-bounded questions are an extension of single- and double-bounded referendum questions. Participants are asked whether or not they would vote to pay for a project over a range of costs. For example, this study used 16 cost levels, ranging from ten cents to $150. Each participant is asked how they would vote at each amount, thus avoiding the anchoring bias problems associated with double-bounded questions.[4] Compared to single-bounded questions, having participants evaluate a wide range of costs may reduce yea-saying since they are not faced with only one chance to support an environmental improvement. Unless the participant answers "yes" to every amount, or "no" to every amount, the multiple-bounded approach will provide fairly tight upper and lowers bounds on each participants' WTP.

The polychotomous-choice format allows respondents to indicate how certain or uncertain they are that they would vote yes or no to any particular amount. For example, the five voting choices offered in this survey were "Definitely Yes", "Probably Yes", "Not Sure", "Probably No", and "Definitely No". Allowing respondents to express uncertainty may provide insight into the valuation processes of respondents. When combined with multiple-bounded questions, the researcher can observe the point on the individual's demand curve where WTP starts to become uncertain. This format seems appropriate for wetlands valuation studies since

[4] There may be biases associated with the range of costs covered and intervals between costs, or with whether listing costs in ascending or descending order affects welfare estimation. Hopefully thorough use of qualitative design tools, such as focus groups, will avoid or reduce these potential biases; however, they certainly warrant further research.

respondents typically will not have had a lot of experience searching the preferences for wetlands. As with other formats that do not estimate point values, MB-PC is less statistically efficient than the open-ended question format. However, since it does provide fairly tight upper and lower bounds for most respondents, it is more efficient that single- or double-bounded dichotomous choice questions. This gain in statistical precision is welcome in any study, but seems particularly important for studies that explicitly seek to test sensitivity to scope.

Using a mail survey, a random sample of 550 Wisconsin residents was asked to consider a hypothetical alternative alignment of the two new lanes that would travel around the existing wetlands instead of through them. The survey specified that the alternative alignment would have the same safety and economic benefits as the planned alignment, but would preserve the original wetlands. Survey participants were also informed that the alternative alignment would cost more to construct, and were asked about their willingness to pay higher state income taxes in order to preserve the original wetlands. After three rounds of mailing and a follow-up postcard, the survey had a response rate of 60%. Removing those who did not properly complete the valuation question left 117 respondents in the base group and 122 in the scope group.

A True-False quiz was included on the inside cover of the questionnaire booklet. This type of quiz serves two purposes. First, it presents respondents with an opportunity to review to review the information provided, which leads to more informed responses. Second, it provides the researcher with quantitative evidence of content validity. The quiz in this survey differed between the two groups. Base group respondents were asked eight questions, six of which asked whether scientists expected the mitigation wetlands to replace each of the six wetland services. The other two questions matched two for the scope group, who answered six other questions as well, many of which focused on the quality of the original wetlands. Overall, the base group

correctly answered 83.9% of the questions, and the scope group answered 88.4% correctly. Of the fourteen questions asked, all but three had correct response rates in excess of 80%, with the lowest being 69%. These results provide encouraging evidence of content validity.

The test of sensitivity to scope was accomplished by splitting the sample into two groups. The base group was informed about which services the mitigation attempt was expected to replace. The scope group, on the other hand, was not informed that a mitigation attempt was being conducted. The two groups were thus asked to value projects that differed in the net services provided (i.e., that differed in scope), since one group stands to have some services replaced by the mitigation attempt, but the other group does not. Economic theory predicts that the scope group's WTP will exceed that of the base group since the scope group stands to lose more wetland services if the alternative alignment is not adopted. The null hypothesis of the scope test, therefore, is that WTP will be equal across groups, and the alternative hypothesis is that the scope group's WTP is greater than the base group's WTP. Given the structure of this scope test, a one-tailed test is appropriate.

Scope Test Results

To conduct the scope test, sample means need to be estimated, and then some method of comparison needs to be determined. In order to estimate sample means from polychotomous-choice data, the data first need to be converted to dichotomous-choice data. For the data in the USH 53 survey, three levels of WTP certainty were used to dichotomize the data. The "Definitely Yes" dichotomizing scheme counts only the responses in column A (see Figure 4.1) as Yes votes, and the rest as No votes. The "Probably Yes" scheme considers responses in columns A or B to be Yes votes, and the rest to be No votes. Finally, the "Not Sure" scheme

counts responses in columns A, B, or C as Yes votes, and all others as No votes. Sample means will, of course, increase as decreasing levels of WTP certainty are allowed to count as Yes votes.[5] The ambiguity over how to interpret answers to polychotomous-choice questions need not be viewed as negative. This value elicitation format allows the researcher to present a range of WTP estimates, and can be interpreted as providing upper and lower bounds on WTP.[6] When a more specific estimate is required, the conservative "Definitely Yes" level is in keeping with the NOAA panel's suggestion that when choices need to be made, CV researchers should lean toward the conservative. As will be seen, the scope test results are robust across certainty levels.

To further explore robustness, we examined three data sets that differed in their handling of potential outliers. Four respondents in the base group, and six in the scope group, answered "Definitely Yes" to every bid offered in the WTP question, including the highest bid of $150. These responses might be considered outliers. They may well represent the individual's true valuation, or they may be a form of protest bid. Determining which respondents are being truthful and which are expressing inflated valuations is, of course, a subjective and essentially arbitrary process. If we accept the principle of consumer sovereignty, then it seems we should accept the answers as being truthful unless we have evidence to the contrary. Such evidence would have to come from answers to other survey questions. The types of evidence that might be indicative of

---

[5] The use of a polychotomous-choice question format is not strictly compatible with the random utility model, which assumes that the individual knows his or her utility (and therefore WTP) with certainty. A more complex statistical model that incorporates uncertainty on the part of the respondent might allow for more of the polychotomous-choice WTP information to be utilized than does the present dichotomization scheme. The handling of WTP uncertainty in the context of CV is clearly an area worthy of further research.

[6] Welsh and Poe (1998) compare the various certainty levels of the MB-PC question to single-bounded, payment card and open-ended question formats for the same environmental amenity, which helps illuminate how respondents deal with uncertainty in these latter formats.

untruthful WTP responses are rejection of key elements of the scenario, skepticism about the valuation exercise, and answers that indicate either a lack of support for the alternative alignment or a lack of interest in the environment. The ten potential outliers were screened with these considerations in mind. Scope test results are presented below for three data sets. The full data set contains all 239 participants; the "no outliers" data set removed all ten potential outliers, and the "best" data set has only three of the outliers removed (two from the base group and one from the scope group) based on answers to other questions that identified them as potential protest bids. As will be seen, scope tests results are robust across all three data sets.

Sample means were estimated using maximum likelihood estimation, with a logistic cumulative density function specified for the WTP distribution. As expected, sample means were higher for the scope group than for the base group. Sample means for the base group were $13.68, $24.07, and $37.38 for the three certainty levels, respectively. Sample means for the scope group were $20.77, $41.84, and $57.83, respectively.

One way to compare these sample means is to use the method of convolutions described by Poe, Severance-Lossin and Welsh (1994). The convolutions approach allows for a semi-parametric comparison of two distributions of WTP estimates. Empirical WTP distributions can be obtained from estimates of the mean and variance-covariance matrices of the estimated coefficients by using resampling techniques, such as Monte Carlo or bootstrapping. The method of convolutions then provides an exact statistical test of the difference between these two empirical distributions. This approach essentially computes a third distribution that is the difference between the two empirical WTP distributions, chops off the tail(s) according to the specified significance level, and sees if the remaining distribution includes zero. If not, the WTP distributions are significantly different.

The Monte Carlo resampling technique was used to generate 1000 estimates of mean WTP for each group. Next, the P-value of the convolution of these two empirical distributions was calculated, and this two-step procedure was repeated the procedure 100 times. Results for all three data sets are presented in Table 1. Looking first at the "best" data, average P-values for the three dichotomizing schemes were .0097, .0005, and .0038, respectively. Thus the null hypothesis of equal WTP across treatments can be rejected at the 1% significance level for all three degrees of WTP certainty.

Table 1: One-Tailed Method of Convolutions Scope Tests

WTP Certainty Level	Full Data	Best Data	No Outliers
Definitely Yes	.0419	.0097	.0597
Probably Yes	.0017	.0005	.0023
Not Sure	.0092	.0038	.0123
N (base, scope)	(117,122)	(115,121)	(113,116)

Recall that the "best" data omitted two "outliers" from the base group, but only one from the scope group. This serves to mildly accentuate the difference between the two groups. When scope tests are run on the full data, average P-values are naturally higher. However, the scope test still rejects the null hypothesis of equality at the 5% significance level for all three WTP certainty levels, and at the 1% level for both the "Probably Yes" and "Not Sure" levels. Thus the conclusion is the same - that WTP estimates from this survey are sensitive to scope.

Similarly, scope tests were performed on the "No Outlier" data, which omitted four observations from the base group and six from the scope group. This data should reduce the difference in WTP between the two groups. Here again, the null hypothesis can be rejected for

the "Definitely Yes" scheme at the 10% significance level, for the "Probably Yes" scheme at the 1% level, and for the "Not Sure" scheme at the 5% level. Even with all ten potential "outliers" removed, the WTP estimates are clearly sensitive to the scope of the change in wetland services.

Likelihood ratio tests could also be used to conduct the scope test. In this context, likelihood ratio tests determine whether the estimated coefficients are equal across samples by comparing value of the log likelihood function (LL) for each the two samples to the value of the log likelihood function for a pooled sample. The likelihood ratio test statistic (LR) is calculated as LR = -2 * [LL(base) + LL(scope) - LL(pooled)], and compared to the chi-square distribution with degrees of freedom equal to the number of estimated parameters. These are two-tailed tests, which allow us to test whether the distribution of WTP is the same across samples, but do not test for the direction of any difference between the samples. In order to compare the statistical power of likelihood ratio tests to that of the convolutions method, a two-tailed convolutions test would be required. Table 2 presents such a comparison, where the results presented in the three columns labeled "LR" are likelihood ratio statistics, and the results in the three columns labeled "P-value" are the average P-values of 100 sets of convolutions (as described above), using a two-tailed truncation scheme.

Table 2: Comparison of Two-Tailed Scope Test Results Between the Method of Convolutions and Likelihood Ratio Statistics

Certainty Level	Full Data		Best Data		No Outlier Data	
	LR	**P-value**	**LR**	**P-value**	**LR**	**P-value**
Definitely Yes	3.65	.0837	7.44**	.0194	3.06	.1194
Probably Yes	11.24***	.0033	15.14***	.0009	11.55***	.0046
Not Sure	6.27**	.0183	8.11**	.0075	5.92*	.0246
N (base, scope)	117, 122	117, 122	115, 121	115, 121	113, 116	113, 116

* Significant at the 10% level; ** Significant at 5%; *** Significant at 1%.

As noted, a one-tail test is more appropriate given the hypothesis we seek to test, but this two-tailed comparison is presented in order to demonstrate that the convolutions approach is statistically more powerful than likelihood ratio tests. Note the three shaded boxes in Table 2. In each of these cases, the convolutions method would reject the null hypothesis at lower standard significance levels than would the likelihood ratio test. For example, using the full data set and looking at the "Definitely Yes" level of WTP certainty, the likelihood ratio test would fail to reject the null hypothesis at the 10% significance level, whereas the method of convolutions would reject the null hypothesis. This could obviously lead to very different and erroneous policy recommendations. Thus if the researcher or analyst wishes to assess differences in estimates of mean WTP (or any empirical distribution), s/he should use the convolutions approach to provide a more powerful test.

Conclusions

This study has shown that CVM studies can demonstrate sensitivity to scope, even for complex, relatively unfamiliar goods that are dominated by nonuse values. This result is robust across certainty levels and across data sets with potential outliers removed. The three factors that contributed to this finding were careful attention to establishing strong content validity, the use of a state-of-the-art value elicitation device, and use of powerful econometric techniques. Given the difficult valuation context, this finding is encouraging for the use of CVM. It is also good news for the valuation of wetlands, which would benefit from a more holistic valuation methodology.

References

Bishop, R.C., P.A. Champ and D.J. Mullarkey. 1995. "Contingent Valuation." In *Handbook of Environmental Economics*. Ed. Daniel W. Bromley. Oxford: Blackwell Publishers. 629-654.

Carson, R.T. 1995. "Contingent Valuation Surveys and Tests of Insensitivity to Scope." Draft manuscript.

Desvousges, W.H., F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson and K.N. Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In "*Contingent Valuation: A Critical Assessment*", J.A. Hausman ed. North-Holland, Amsterdam.

Diamond, P.A., J.A. Hausman, F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson, and K.N. Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In "*Contingent Valuation: A Critical Assessment*", J.A. Hausman ed. North-Holland, Amsterdam.

Hanemann, W.M. 1994. "Valuing the Environment Through Contingent Valuation." *Journal of Economic Perspectives.* 8(4):19-43.

Kahneman, D. and J.L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management.* 22(1):57-70.

Mullarkey, D.J. 1997. "Contingent Valuation of Wetlands: Testing Sensitivity to Scope." Ph.D. dissertation, University of Wisconsin, Madison.

Poe, G.L., E.K. Severance-Lossin and M.P. Welsh. 1994. "Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics.* 76:904-915.

Schkade, D.A. and J.W. Payne. 1994. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation." *Journal of Environmental Economics and Management.* 26:88-109.

Smith, V.K. 1992. "Comment: Arbitrary Values, Good Causes, and Premature Verdicts." *Journal of Environmental Economics and Management*. 22(1):71-89.

Welsh, M.P. and R.C. Bishop. 1993. "Multiple Bounded Discrete Choice Models." In J.C. Bergstrom, ed., *Benefits & Costs Transfer in Natural Resource Planning.* Western Regional Research Publication, W-133, Sixth Interim Report. 331-352.

Welsh, M.P. and G.L. Poe. 1998. "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." *Journal of Environmental Economics and Management*. 36:170-185.