ESO 2277

A BENEFIT EVALUATION OF PROGRAMS TO ENHANCE GROUNDWATER QUALITY, SURFACE WATER QUALITY AND WETLAND HABITAT IN NORTHWEST OHIO

Damitha de Zoysa, Alan Randall and Fred Hitzhusen*

A. Introduction

The overall objective of this study was to design a comprehensive split-sample contingent valuation (CV) instrument that would allow the estimation of benefits of three environmental services, enhancements to groundwater, surface water and wetland habitat. Two other components of the overall objective that emerged from this study are:

1. Develop a conceptual framework for incorporating multi-dimensional programs to generate benefits of improving/protecting surface water, groundwater and wetland habitat services using a Total Value (TV) framework.
2. Test empirical hypotheses using the multivariate analysis of the relationship between yes responses to the offered program and a set of explanatory variables.

The first sub-objective, i.e., developing a conceptual framework for incorporating multi-part policies, has been addressed. Most of the conceptual and practical problems with using conventional Independent Piecewise Valuation (IPV) and valid design Sequential Piecewise Valuation (SPV) methods using a TV framework have been articulated. Based on the theoretical and empirical progress accomplished in valuing

---

*Graduate Research Associate and professors, respectively, Department of Agricultural Economics, The Ohio State University, Columbus, OH 43210, December 1995.

multi-part policies (Hoehn and Randall, 1989, Hoehn, 1991, Hoehn and Loomis, 1993), a TV framework for developing valid methods of using independent estimates to value multi-part policies in conducting benefits transfer (BT) is discussed. However, no attempt was made to use the generated benefits in conducting BT experiments in this research.

B. Summary of the Research Findings

The experimental design adopted in the CV survey is an application of split-sample design, as suggested by the National Oceanic Atmospheric Administration (NOAA) Panel. The study has focused on measuring total values, including nonuse values of three environmental resources using the split-sample design to allow for hypothesis testing concerning how willingness to pay is influenced by programs offered and sub-populations sampled. The study focuses on measuring total values (use and nonuse) of three environmental services by evaluating surface water, groundwater and wetland habitat programs in northwest Ohio. We obtain total values for the offered programs from Maumee rural, urban Maumee and Columbus/ Cleveland samples. Because Columbus/Cleveland is an out-of-region sample, nonuse (passive) may be a larger component of TV.

This study applies Distribution With Equal Area Bid Selection (DWEABS) procedure in generating the efficient set of offer prices for the final survey. The DWEABS method allocates the prices among the sample so as to minimize the variance of the WTP welfare measure.

The common practice in analyzing CV data has been to exclude protest bids

from the final data set. However, we took a less conventional approach that paid more attention to the data itself. Using a sequence of questions defined in the CV study instrument (Questions 6, 7 and 8 in the survey), we allowed protesters to self-identify and then performed multivariate analyses first from yes voters and then to see if they differed from no voters. Since there was little obvious difference between protest voters and no voters, we performed all subsequent analyses with YN and YNP data.

Multivariate analysis of construct validity was conducted using probit choice function and maximum likelihood estimation procedures. Results obtained are consistent with prior expectations with a few exceptions related to significance of the estimates. Price variable (combined with sample dummies to permit the price response to differ across samples) is the most significant predictor of the respondents' voting behavior. Income coefficient was positive and significant. Most of the attitudinal variables priority ranking on groundwater program ranging from high priority to low priority(GW1), priority on wetland habitat program ranging from high priority to low priority (WHP1), government spending on education, health and vocational training in terms of more, same and less spending (EHV1) indicated that high priority and spending more on public goods/policies was positively correlated with vote and statistically significant. All the program dummy variables had the expected sign compared to the omitted program confirming monotonicity i.e., that more or same is preferred to less.

One of the objectives of the empirical analysis was to test if the single-

programs were less attractive than two-program agendas and two-program agendas were less attractive than three-program agendas at given prices. The empirical results for YN data set indicated that of 9 tests of WTP conducted, 7 were consistent with prior expectations. Two tests revealed inconsistent results compared to the theoretical norm. The expected median WTP for program six (SWWI - two-program agenda) indicated relatively smaller values in magnitudes ($ 64.81 per household per year) compared to $ 167.98 per household per year for program two (SW-single program). Similarly, the expected median WTP for program six (SWWI) indicated relatively smaller value ($ 64.81 per household per year) compared to $ 72.67 per household per year for program three (WI - single program). This unexpected result might be explained by unusual high non-responses (76 and 80 percent respectively) for program six (SWWI) at higher price levels ($ 80 and $ 120). For YNP data, sign of coefficients for all 9 programs were consistent with theoretical expectations.

The median WTP estimates vary considerably across the samples. The lowest edian WTP is observed for sample 2 and sample 3 in YNP and YN data respectively. Comparison of median estimates for YN and YNP data by sample shows that median WTP values for YNP decrease by 79 percent, 76 percent and 38 percent over YN data for samples 1, 2 and 3 respectively.

Median WTP for sample 3 in YN data showed the lowest values as expected. Although one would expect the lowest WTP values from the out-of-region sample, it is not the case with respect to YNP data in this study. The median WTP using YNP data showed that the highest WTP was from the out-of-region sample ($52.45), while

the lower values came from samples 1 and 2. One would expect lower WTP for the out-of-region sub-sample. It seems likely that this result reflects the concept of median and the nature of the estimated equation (in which sample 3 had a higher intercept and a steeper price slope). If this conjecture holds, the results may be reversed for mean WTP, with sample 3 lower than the other sub-samples.

This study has demonstrated the estimation of mean WTP using a more conservative measure, i.e., Lower-Bound Mean (LBM). Implausible means generated by log-normal distribution beyond a given range of data led to the estimation of LBM using the distribution of estimated Pr(Yes) responses. However, the procedure used in estimating LBM method determines that more of the higher probability distribution (higher valued program) is truncated than for lower distribution (lower valued program). Therefore, using LBM method for comparison purposes would compress the difference between the LBM.

C. Indicators of Validity and Reliability of the Study

Accuracy of the results obtained from our CV survey has been evaluated using the guidelines suggested in the NOAA Panel. Two dimensions of accuracy relate to validity and reliability. However, validity could be judged using both construct and convergent validity criteria. Construct validity can be evaluated by examining whether parameter estimates conform to theoretical expectations. Demonstration of convergent validity is not possible in this study since it requires evaluation of estimates derived from two different elicitation methods. Reliability defined by the Panel is used in a statistical sense which implies that parameter estimates should be statistically reliable

when extrapolated to populations without significant bias with practical sample sizes (McFadden et al. 1992).

In order to test the validity of the CV study, two common areas suggested in the literature; theoretical validity and statistical reliability have been examined (McFadden, 1992).

In order to demonstrate the presence of theoretical validity the results of the construct validity analysis are discussed. We find that wealthy households are more likely to vote for resource preservation/improvement programs than poor households. Price variable has a negative significant effect on the voting behavior as expected. We observe that vote for the environmental programs increases among households who believe that the government should give high priority for such programs. The same observation can be made with regard to spending on education, healthcare and vocational training programs. Respondents who believe that the government should spend more money on education, healthcare and vocational training are more likely to vote for these programs.

One approach to testing the theoretical consistency of CV results is the evaluation of sensitivity of WTP estimates to different levels of natural resource services (programs). In the context of this study it was one of the objectives to examine if the WTP estimates are affected by magnitude of the policy offered, i.e; single program, two-programs and three-programs. In essence the hypothesis assumes that WTP estimates conform to the basic micro-economic principle of monotonicity which tests the theoretical validity of CV surveys. In order to test the hypothesis that

WTP does not increase as programs are added to the agenda, statistical tests were performed for joint and pairwise comparisons. The findings are given in Table 26 for YN data while Table 27 provides results of YNP data (provided in the dissertation).

Except in 2 cases, all our results show that the WTP estimates for the programs are not affected significantly by its magnitude (Table 24). Of 9 combinations, 7 pairs had the expected sign while SW combined with WI (SW > SWWI; WI > SWWI) had the unexpected sign. In terms of parameter statistical significance none of these was significant. Statistical tests performed on pairwise and joint comparisons were not quite significant given small sample size. Results in Table 26 reveal that GWSW (Two-program) is more valuable to GW (single-program) alone. Further, joint test of the programs also suggests that significance is barely rejected ($p = 0.12$) at the 95 percent level of significance. Given the mixture of significance in intercept dummies for programs (Vi) median WTP differences are inevitable. For YNP data, however, all 9 combinations had expected sign as expected. Wald Statistic performed for pair-wise and joint treatments indicated the presence of significant differences between treatments in 3 combinations. "Perfect embedding" can be defined as a situation where the stated WTP is the same for preserving or improving single-program, two-program or three-program agendas. With the presence of perfect embedding we would have found that improving groundwater quality has a value of some magnitude when valued as a single program and of the same value when combined with a surface water program implying that by adding SW program, the value added by the respondent is zero. Results reported in

Tables 35 and 36 show that stated WTP for sum of two independent programs are not equal to the stated WTP values of the combined program.

Our results fail Diamond's (1992) "adding-up" test which asserts that the sum of the values of the independent programs should be equal to the value of the joint programs containing them. However, there is ample theoretical and empirical evidence that adding-up test itself is suspect (Hoehn and Randall, 1989, Hoehn, 1991, Hoehn and Loomis, 1993, Randall and Hoehn, 1996).

Reliability also refers to the hypothesis that DC-CV results should yield statistically defensible comparable estimates to different estimations or analytical procedures. Testing the reliability of DC-WTP estimates can be evaluated using robustness of parameter estimates. We achieve robustness of coefficients in this study implying that parameter estimates are not sensitive to changes in specification. The importance of functional form in estimating WTP using DC has been a well-discussed aspect in CV literature (McFadden, 1992, 1994, Bowker and Stoll, 1988). This study has employed two functional forms; linear-normal and log-normal. However, it was decided to report the estimates using log-normal specification as it provides a better fit to the data in the range of prices offered. Plotting estimated probabilities of yes responses against bid levels suggested that linear-normal forces the tail of the distribution to fall thereby distorting the mean WTP values.

D. Implications

a. Implications for CV Research

Based on our results, we find that overall, the CV study achieves theoretical

validity and reliability in valuing three natural resources.

Our study recognizes the importance of a well-designed questionnaire that best informs the respondents of complex multi-part programs within the institutional framework of the NOAA Panel guidelines. Scientific sampling procedures are also pivotal to valid representation of the population. The aggregate WTP benefits for the population would be accurate if the sample is a good representation of the general population.

The study also suggests the importance of obtaining a range of reasonable prices (DWEABS procedure) that reflect respondent's WTP to some extent. However, efficiency and accuracy of DWEABS procedure cannot be validated for this study unless a different approach is used to obtain a range of prices for the DC data.

Success of the CV survey requires well-defined procedures pertaining to the implementation of the survey. This study relies on a set of procedures articulated by the Total Design Method (TDM) closely in implementation of the survey.

The results of this study indicates that the choice of functional form needs careful consideration in evaluating respondents' WTP to the offered programs. We find that the logarithmic specification is superior to linear specification within range of prices. However, logarithmic functional form also leads to implausible means beyond the range of data given. One possible solution would be to consider more flexible functional forms in accommodating multi-part programs and split-samples simultaneously (systems of equations with higher order terms to take into account effects of substitution and complementarity between programs). As a conservative

measure we have estimated truncated means (LBM) using the estimated probabilities of yes responses instead of expected mean WTP.

One of the purposes of the stratified sample design is to observe how voting behavior differs across in-region and out-of-region samples. We conclude that even with an out-of-region sample that might be expected to have a large nonuse component of total value, it is misleading to label their WTP values solely being attributed to nonuse. All we can conjecture is that being an out-of-region sample they may have a higher component of nonuse values in their TV estimates. Separating use from nonuse values generated in CV surveys needs careful interpretation (unless the components of nonuse values are specified in the survey). Thus, on the explanation of behavior of out-of-region sample in YNP data we find substantial contribution of nonuse values in evaluating non-market goods in a total value framework. The benefits estimates would have been underestimated if the WTP of out-of-region citizens were ignored.

We also demonstrate a different way to accommodate protest voters into the analysis by deviating from the conventional CV method of treating protesters.

Comparison of WTP estimates across programs allowed us to examine the much discussed embedding effect in CV findings. The empirical analysis conducted in examining the embedding effect was facilitated by the design of multi-programs of the instrument. This suggests that a more comprehensive, methodological design of program agendas is needed for a thorough analysis of embedding effects.

Finally, selection of a fair and equitable payment vehicle is of utmost

importance for the successful implementation of any CV instrument. We find that designing a payment vehicle for multi-county but sub-state is challenging.

b. Implications for Environmental Policy

CV methods have been recognized as the most promising tool to obtain useful welfare measurements of environmental resources (in the absence of a well-defined market) in formulating public environmental policies. Ability to capture the benefits of nonusers of the resource has been the unique feature in CV method.

Results obtained in this study provide benefits information to public policy makers, as well as a perspective on the preferences of the public concerning providing incentives to farmers to reduce chemical usage and protect resources from deterioration. These estimated benefits could be used to compare benefits of such multi-programs or single programs across the states. The study also would be able to provide some useful information in design of multi-part programs that combine three different resources.

The benefit analysis of surface water, groundwater and wetland habitat improvement/protection could serve as libraryshelf information needed in conducting benefits transfer experiments using the demand function. Even in Ohio itself these benefits can be used for inter-region (between) and intra-region (within) comparisons.

Benefit estimates for 7 programs indicate that residents in northwest Ohio are willing to pay for improvement cum protection of surface water, groundwater and wetland habitat. Comparison of WTP (costs to the residents) indicates the relative spending ranking in terms of public goal policies for environmental services. The

benefits of abatement of contamination of water quality and wetland habitat preservation vary across samples. The benefit measures of 7 different programs also suggest the need for considering implementation of joint programs instead of single programs in terms of benefit measures obtained. The substantial WTP obtained for preserving wetland habitat program shows the high value of programs to convert unproductive marginal farm lands for the purpose of restoring wetlands.

Aggregate benefits are reported as one-time payments per acre of GW, SW and WI respectively. We report a large array of aggregate benefit ($) estimates using both median and LBM measures for different populations (Tables 40-43). Of these aggregate benefits we chose LBM estimates computed for YNP data based on the assumptions discussed previously (Table 43). For GW program aggregate benefits per acre of cropland are $ 4.04 for Maumee rural and urban Maumee combined. Per acre benefits for SW program is a little higher compared to GW program and is $ 6.05 for Maumee rural and urban Maumee combined. Similarly for MR+UM+CC benefits per acre of cropland for GW and SW are $ 17.55 and $26.06 respectively. For the WI program, benefits per acre of wetland are $ 85,214.97 when aggregated across the whole population of Ohio compared to $ 21,566.48 for MR+UM+CC.

E. Limitations of the Study

Often no follow-ups on non-respondents are observed in CV survey. We did not attempt to follow-up on non-responses due to time and budget constraint. However, a follow-up on non-responses would have been desirable. Non-responses have become an important concern in CV surveys because a reasonable response rate

of the representative sample is essential in order to draw accurate inferences about a population.

Before correcting for the problem of non-responses, it is important to recognize the magnitude of the problem. Non-response bias can be a major problem if the response rates are low, i.e., 50 percent or less is considered to be a low response rate in mail surveys (Dillman, 1978). In our study we obtained 58 percent response rate for the target area (Maumee rural) and overall response rate for the entire sample is 50 percent. It would be naive to assume that 50 percent response rate suggests no non-response bias in the analysis. Yet, comparing the response rate obtained in this study with that of similar CV studies (most of them have not rectified the issue of non-responses with relatively smaller response rates), we find no substantial problems with our estimates due to presence of non-response bias. In mail surveys, sample selection bias or self-selection bias may be possible for two reasons (Heckman, 1979). First, self-selection bias may arise due to respondents' choice to respond to the survey compared to those who do not. Second, researcher's influence on the sample selection process may affect the sample selection bias. Given the response rate for CV surveys of this nature, the presence of non-response bias may be minimum.

A more complete and accurate analysis on WTP on cells which combine samples and versions for both data sets would be possible with the increased number of responses. The same justification is applicable in deriving more plausible means given a larger data set. One other problem faced (as a result of a thin data base) was the inability to estimate the Turnbull Lower-Bound Mean (TLBM is a non-parametric

method) using actual probabilities of yes responses due to inadequate data. Once we controlled for socio-demographic variables in the model we were able to estimate the lower-bound mean of the expected WTP instead of a expected mean WTP.

F. Future Research

One area that was not addressed in this study was the analysis of the maximum WTP data. That would allow some gain in efficiency in estimating WTP values by incorporating more information on voting behavior. Moreover, the design of the survey instrument could also be used in analyzing the substitution and complementarity effects between programs (Hoehn and Loomis, 1993). There is also a need to establish a comprehensive method of dealing with C.I. for sample and versions as well as cell data.

The split-sample design allows for benefits transfer experiments using the estimated WTP function. The information on WTP can be used in transferring benefits to evaluate similar programs in other states if the socio-demographic conditions are similar. However, a more accurate way would be to estimate the benefit function or valuation function by making some modifications to the analysis. Inter and intra regional comparison of WTP and validation of benefits estimates using statistical tests would be of some interest to the researcher. As a final evaluation procedure, technical criteria could be used for evaluating transferability of benefit estimates to the policy site.

This study has demonstrated the benefits of contamination abatement (benefits of environmental programs to the consumers) for three services. In order to conduct a

benefit - cost analysis of environmental resources, cost of abatement needs to be identified. Cost side of the analysis requires focusing on on-farm economic effects of strategies to reduce contamination problem using a bio-physical and bio-economic model. An accurate economic assessment of the biophysical process concerning leachate requires input from several disciplines (e.g., engineering, geohydrology and plant science etc.). The assessment reflects linkage between producer behavior (management practices) and groundwater, surface water and wetland pollution by integrating hydrologic and biophysical models of farm level processes.

An obvious extension is to do a benefit-cost analysis which would require substantial information on costs in delivering these programs. However, a sub-set of information may be obtained from Toledo Ports Authority/Maumee River Basin project in the case of surface water program.

References

Bowker, J.M. and J.M. Stoll. 1988. "Dichotomous Choice in Resource Valuation." American Journal of Agricultural Economics, Vol. 70, pp. 372-82.

Diamond, P.A. et al. 1992. "Does Contingent Valuation Measure Preferences? Empirical Evidence." Cambridge Economics Inc., Washington, D.C.

Dillman, Don A. 1978. "Mail and Telephone Surveys: The Total Design Methods." Washington State University, Pullman, Washington

Heckman, James J. 1979. "Sample Selection Bias as a Specificationn Error." Econometrica, Vol. 47, No. 1, pp. 153-161.

Hoehn, John P. 1991. "Valuing the Multidimensional Impacts of Environmental Policy: Theory and Methods." American Journal of Agricultural Economics, Vol. 73, pp. 289-99.

Hoehn, John and Alan Randall. 1989. "Too Many Proposals Pass the Benefit Cost Test." American Economic Review, Vol. 79, pp. 543-51.

Hoehn, John P. and John B. Loomis. 1993. "Substitution Effects in the Valuation of Multiple Environmental Programs." Journal of Environmental Economics and Management, Vol. 25, pp. 56-75.

McFadden, Daniel. 1994. "Contingent Valuation and Social Choice." American Journal of Agricultural Economics, Vol. 76, pp. 689-708.

McFadden, Daniel and Gregory K. Leonard. 1992. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." in Cambridge Economics Inc., pp. 2-63.

Randall, A., and John P. Hoehn. 1996. "Embedding in Market Demand Systems." Journal of Environmental Economics and Management, Vol 30(1). in press.