NA

Survey Response-Related Biases in Contingent Valuation: Concepts, Remedies, and Empirical Application to Valuing Aquatic Plant Management

Mark L. Messonnier, John C. Bergstrom, Christopher M. Cornwell, R. Jeff Teasley, and H. Ken Cordell

Sample nonresponse and selection biases that may occur in survey research such as contingent valuation applications are discussed and tested. Correction mechanisms for these types of biases are demonstrated. Results indicate the importance of testing and correcting for unit and item nonresponse bias in contingent valuation survey data. When sample nonresponse and selection bias go uncorrected, welfare measures may be overestimated or underestimated contributing to potential errors in resource policy and management decisions.

Key words: survey research, contingent valuation, sample nonresponse bias, sample selection bias

Techniques used in the valuation of environmental amenities and natural resources often rely on surveys as a means of collecting data. Whatever the strengths and weaknesses of the various valuation techniques, the survey process may introduce problems of its own. Among the most troubling of these problems, especially as it affects mail-administered surveys, is the potential for the introduction of what is commonly referred to as nonresponse bias or sample bias. These problems we refer to as survey response-related biases (SRB). The term sample bias is not specific enough to communicate the nature of the problem, while the term nonresponse bias is so specific that it implies that nonresponse is the general cause of the bias. As we discuss below, random failure on the part of survey questionnaire recipients to respond is a passive potential source of bias. Systematic, interest-induced responses to survey questionnaires are potential active sources of bias.

Several researchers have recently examined the problem of SRB in the contingent valuation (CV) willingness-to-pay (WTP) setting. A general overview of sampling issues in contingent valuation, including* nonresponse bias, is provided by Mitchell and Carson. Edwards and Anderson found significant differences between characteristics of respondents and nonrespondents. They also found no significant bias introduced by culling outliers, though this result is somewhat unsatisfying. Intuitively, one would expect self-selection to have far more influence on the size of the sample than would culling. Whitehead, Groothuis, and Blomquist found no significant difference between respondents and norespondents in their sample. They used data from a previous survey of the same sample to impute values for missing elements. No indication was given as to why the data from the previous survey were not used instead of imputation, and they discussed a correction technique for only one form of SRB. Another application of testing for SRB is reported in Whitehead (1991). The response rate from a general sample was compared with the response rate from an interest group sample. Mem-

---

Mark L. Messonnier is a doctoral graduate, Department of Agricultural and Applied Economics, University of Georgia, Athens; John C. Bergstrom is a professor, Department of Agricultural and Applied Economics, University of Georgia, Athens; Christopher M. Cornwell is an associate professor, Department of Economics, University of Georgia, Athens; R. Jeff Teasley is a research coordinator, Department of Agricultural and Applied Economics, University of Georgia, Athens; and H. Ken Cordell is a research social scientist and project leader, Southern Research Station, U.S. Department of Agriculture Forest Service, Athens, GA.

Funding for this project was provided by the Georgia Agricultural Experiment Station, the USDA Forest Service, the U.S. Army Corps of Engineers, and the Tennessee Valley Authority.

bers of the interest group were found to hold significantly higher values and responded to the survey in significantly greater proportion than did the general sample. Correction methods were briefly discussed but not used. Whitehead et al. (1994) defined a sample—from a study using continuous WTP responses—as the population and then simulated response rates to test for SRB. They weighted the data to make it representative and followed Heckman in estimation to correct for biases. Results showed SRB to be significant at both low and high response rates.

Loomis examined the bias resulting from failure of survey recipients to return the instrument. Several approaches to correction were discussed and demonstrated, although none of the measures are adequate to address how nonresponse may systematically affect WTP. Whitehead (1994) used imputation methods to correct for bias resulting from returned but incomplete survey instruments. However, the use of imputation techniques creates its own set of inference problems and still does not account for correlation between determinants of response and WTP.

These studies indicate a heightened awareness of SRB. However, what is lacking in the current applied welfare economics literature is a single work making a clear presentation of the concepts and definitions involved in SRB, the reasons why it can result in estimation bias, and how it can be empirically treated in CV studies that use either continuous or dichotomous WTP responses. Our objective is to help fill this literature gap by defining SRB and its two elements, sample nonresponse bias (SNB) and sample selection bias (SSB), discussing the common correction methods, and then developing a general model of SSB in CV. A special case of the general model is developed to highlight the potential sources of SSB. As an illustration of the use of the techniques, we apply them to a natural resource valuation problem. Mail survey data are used to estimate the recreational use value of a northeastern Alabama reservoir under several management alternatives. Estimates are found with and without corrective measures taken. We intend the discussions and models to be helpful to CV practitioners in their design of surveys and estimation of welfare benefits.

Sample Nonresponse Bias

If all members of the sample to whom questionnaires are mailed-the surveyed sample-respond with fully complete questionnaires, then there is no possibility of SRB. However, it is unlikely that all recipients will respond, so that the realized sample (the final data set available to the researcher) will differ from the surveyed sample. Consequently, the sample available for estimation may not be representative of the target population, a clearly undesirable circumstance. There are two reasons for the difference. A recipient may simply not return the questionnaire, thereby exhibiting unit nonresponse. No information is available on that member of the surveyed sample. When the recipient returns the questionnaire but does not answer certain questions of importance to the study, item nonresponse has occurred. As long as other questions are answered, some information is available on item nonrespondents. A survey can suffer from both types of nonresponse, both of which remove an observation from the realized sample.

Nonresponse may occur randomly in the sense that the factors affecting the probability of response are not correlated with the factors affecting WTP. If, as a result of this kind of random nonresponse (either at the unit or item level), identifiable subgroups of the population are disproportionately represented in the sample used for estimation, and these subgroups hold WTP values that differ from other subgroups in the population, then SNB exists. SNB is a "passive" concept. Some questionnaire recipients simply do not respond, and there is no systematic relationship between their nonresponse and their WTP. The nonresponse, however, may lead to a realized sample that is not representative of the target population. Estimation of a WTP function with nonrepresentative observations will result in biased inference with respect to population parameters. The error will be further compounded by evaluation of the estimated WTP function with biased sample statistics. The essence of SNB is that the sample cannot speak directly to the population.

Solutions to the problem vary with its source. To correct for SNB due to unit nonresponse, the available complete responses may be weighted so that the weighted sample statistics correspond to population parameters obtained from census data (Anderson, Basilevsky, and Hum). Somewhat different SNB correction techniques are used for item nonresponse. Because there is some information available on item nonrespondents, it can be used to impute missing WTP values (David et al.). Observations missing WTP or other values can be assigned those values on the basis of their similarity to other observations with like characteristics. Alternatively, WTP values may be obtained through regression (Orchard and Woodbury, Dempster, Laird, and Rubin). While they serve to reduce bias, imputation techniques result in increased variance and so make inference about particular determinants of WTP difficult (Mitchell and Carson).

Sample Selection Bias

Unit and item nonresponse may occur in such a way that there is correlation between the factors that determine survey response and the fautors that affect WTP When this occurs, SSB has its effect through the error term of the WTP equation. SSB is an "active" concept. Survey recipients who have a greater interest in the commodity of interest are more likely to respond and thus will be over-represented in the realized sample. They may hold either higher or lower values for the commodity of interest than other members of the population. As in the case of SNB, it is useful to examine response rates of subgroups, but here interest lies in the response rates within those subgroups rather than between them. The question to be asked is whether there is something systematically working to cause some members of a particular subgroup (e.g., surveyed sample) to respond while other members do not.

Heckman characterized the problem as one of misspecification due to an omitted variable. To see this, let the econometric representation of the SSB issue be given by the set of equations

$$(1)\qquad \mathrm{WTP}_i = \beta'_{\mathrm{WTP}} X_{\mathrm{WTP},\,i} + \varepsilon_{\mathrm{WTP},\,i} \qquad i = 1, \ldots, n$$
$$\mathrm{RESP}^*_i = \beta'_{\mathrm{RESP}} X_{\mathrm{RESP},\,i} + \varepsilon_{\mathrm{RESP},\,i} \qquad i = 1, \ldots, N$$

where WTP_i is the respondent's willingness-to-pay, RESP^*_i is the latent propensity to respond, $X_{\mathrm{WTP},\,i}$ and $X_{\mathrm{RESP},\,i}$ are vectors of explanatory variables for the WTP and response equations, and $\varepsilon_{\mathrm{WTP},\,i}$ and $\varepsilon_{\mathrm{RESP},\,i}$ are identically independently distributed (i.i.d.) normal errors. The observable counterpart to RESP^*_i is RESP, which takes on a value of 1 when RESP^*_i is nonnegative, and 0 otherwise. The number of unit respondents is denoted by N. The number of complete, usable responses (the number of item respondents or the realized sample) is denoted by n, with the number of item nonrespondents given by $(N - n)$. The WTP equation is called the primary equation while the second is referred to as the selection equation or the sample selection rule (SSR), since it is the rule by which unit respondents self-select into the sample of n complete observations.

Due to the failure of some unit respondents to provide valid data on either WTP or one or more of the values in X_{WTP}, the primary equation can only be estimated with the realized sample. Therefore, the regression of the primary equation over the realized sample takes the form

$$(2)\qquad E(\mathrm{WTP}_i \,|\, X_{\mathrm{WTP},\,i}, \mathrm{SSR}) = \beta'_{\mathrm{WTP}} X_{\mathrm{WTP},\,i} + E(\varepsilon_{\mathrm{WTP},\,i} \,|\, \mathbf{X}_{\mathrm{WTP},\,i}, \mathrm{SSR}) \qquad i = 1, \ldots, n.$$

Suppose the SSR is that WTP_i is observed if RESP_i is equal to one. If $\varepsilon_{\mathrm{WTP},\,i}$ and $\varepsilon_{\mathrm{RESP},\,i}$ are independent, then

$$(3)\qquad E(\varepsilon_{\mathrm{WTP},\,i} \,|\, X_{\mathrm{WTP},\,i}, SSR) = E(\varepsilon_{\mathrm{WTP},\,i} \,|\, X_{\mathrm{WTP},\,i}, \mathrm{RESP}_i \geq 0) = E(\varepsilon_{\mathrm{WTP},\,i}\, X_{\mathrm{WTP},\,i}, \varepsilon_{\mathrm{RESP},\,i} \geq -\beta'_{\mathrm{RESP}} X_{\mathrm{RESP},\,i}) = 0.$$

Hence, the factors affecting response are not correlated with the magnitude of WTP. On the other hand, if $E(\varepsilon_{\mathrm{WTP},\,i} \,|\, X_{\mathrm{WTP},\,i} \mathrm{SSR}) \neq 0$, the regression for the realized sample is

$$(4)\qquad E(\mathrm{WTP}_i \,|\, X_{\mathrm{WTP},\,i}, \mathrm{RESP}_i \geq 0) = \beta'_{\mathrm{WTP}} X_{\mathrm{WTP},\,i} + E(\varepsilon_{\mathrm{WTP},\,i} \,|\, \varepsilon_{\mathrm{RESP},\,i} \geq -\beta'_{\mathrm{RESP}} X_{\mathrm{RESP},\,i})$$

and is clearly dependent on $X_{\mathrm{WTP},\,i}$ and $X_{\mathrm{RESP},\,i}$. Estimating the primary equation in (1) when (4) is the true model generates SSB the source of which is an omitted variables problem.

Heckman provides a simple two-step approach to detection and resolution of SSB that is widely known and accepted. Heckman's original work focused on a model in which the primary equation's regression was continuous and thus estimable by ordinary least squares (Davidson and MacKinnon). Most current CV studies, however, use the closed-ended or dichotomous-choice (DC) question format. A Heckman two-stage (H2S) approach in the DC WTP equation context would involve estimating the conditional probability

$$(5)\quad \text{Prob}[\text{WTP}_i = 1|\text{RESP}_i = 1] = \Phi(\beta'_{\text{WTP}} X_{\text{WTP},\,i} + \beta_\lambda \lambda_i)$$

where Φ is the standard normal cumulative density function (cdf) and λ_i is the inverse Mills ratio representing the probability of person i being in the realized sample.

Equation (1) illustrates a CV survey situation in which there is a single, general source of SSB. However, SSB can result from unit nonresponse and(or) item nonresponse. In a DC WTP elicitation format, both types of nonresponse can be considered simultaneously as a special case of (1) in the system of equations

$$(6)\quad \begin{aligned} \text{WTP}^*_i &= \beta'_{\text{WTP}} X_{\text{WTP},\,i} + \varepsilon_{\text{WTP},\,i} \\ & i = 1, \ldots, n_r \\ \text{ITEM}^*_i &= \beta'_{\text{ITEM}} X_{\text{ITEM},\,i} + \varepsilon_{\text{ITEM},\,i} \\ & i = 1, \ldots, n_u \\ \text{UNIT}^*_i &= \beta'_{\text{UNIT}} X_{\text{UNIT},\,i} + \varepsilon_{\text{UNIT},\,i} \\ & i = 1, \ldots, N_s. \end{aligned}$$

WTP^*_i, ITEM^*_i, and UNIT^*_i are unobserved, but WTP_i, ITEM, , and UNIT_i are observed, taking on a value of one if their unobserved counterparts are nonnegative, and zero otherwise. The error terms are each normally distributed with zero means and variances normalized to one. The size of the surveyed sample is given by N_s, the number of unit respondents by n_u, and the number of item respondents-the realized sample-by n_r. The selection equations are reduced forms and should therefore be viewed as predictors of response rather than having important economic content and interpretation.

All recipients of a questionnaire must decide whether to return it or not, and so the unit selection equation applies to all N_s members of the surveyed sample (if returned, $\text{UNIT}_i = 1$). Of this number, n_u make the decision to return the questionnaire and therefore must make the item response decision. The item selection equation is estimated over the n_u unit respondents. The realized sample is composed of those recipients who return the questionnaire and also provide answers to the DC WTP question and the questions about the determinants of WTP (if all of these questions are answered, $\text{ITEM}_i = 1$). These are the n_r observations which can be used to estimate the primary equation.

The model can be considered in a trivariate specification in which both sources of selection bias, unit and item nonresponse, are analyzed jointly with the primary equation. In principle, a trivariate probit treatment would serve in a joint analysis of three dichotomous outcome variables. As a practical matter, however, the obstacle to this approach is the difficulty in evaluating multivariate normal integrals of order higher than two. An alternative is to proceed in a H2S fashion analogous to (5) above. The two selection equations can be estimated separately and variables can then be constructed to represent item and unit nonresponse bias for inclusion in a univariate probit estimation of the primary equation, as in (5).

Another way of addressing the problem is to model the WTP equation and each selection equation in a pairwise fashion, that is, to specify a WTP/item model and a WTP/unit model. The sources of SSB can then be analyzed separately. This can be done in a H2S fashion as shown above. In the case of item nonresponse, the data for applying this approach are readily available from the returned questionnaires. However, unit response may be more of a problem because of data limitations. The researcher typically has little or no access to data describing the unit nonrespondents. As shown in the empirical application below, if the researcher can obtain data on unit nonrespondents, the H2S methodology can be applied to test for and correct unit and item nonresponse bias.

Empirical Application

As part of a larger study of aquatic plant management at Tennessee Valley Authority (TVA) and U.S. Army Corps of Engineers (USACE) reservoirs, a CV study was conducted to estimate annual recreational use values at Lake Guntersville, Alabama under varying aquatic plant control alternatives. Lake managers were interested in how different types of users value plant management. Surveyed users were therefore classified as fishers and nonfishers according to their self-reported primary use of the lake. The data collection process was intentionally designed so that the effects of SRB could be accounted for in estimation. Questions relevant to the benefit estimation portion of the project were included in two pre-CV surveys to provide backup data in the event that CV response would be less than complete. Copies of the survey instruments along with their cover and reminder letters are available from the authors upon request. These surveys are described below.

On-Site Survey

The on-site use survey collected data on recreation use at Lake Guntersville such as length of stay, number of annual trips, and main activity. The on-site survey was administered at various recreational access points surrounding Lake Guntersville. The form was given in a face-to-face interview by one of two survey team members during the hours of 9:00 a.m. to 5:00 p.m. Users of the lake entering and exiting outside of these hours were not surveyed but were accounted for in the use estimation through the use of traffic counters. The survey sites corresponded to a stratified random sampling plan, except in cases of severe weather or site obstruction.

In developing the strata for the sampling plan, emphasis was placed upon identifying key variables that might account for potential variation within the sample. It was determined that there are different types and amounts of use occurring in different parts of the lake during the different seasons throughout the year. The strata types were defined by ***season*** (fall/winter, spring, or summer), ***day type*** (weekend/holiday, or weekday), ***geographical zone*** (northern, middle, and southern sections of the lake), and ***site type*** (designated developed recreation areas, boat launching and service facilities, and dispersed and informal recreation areas).

The primary sampling plan consisted of fifty-four strata using the above definitions. The basic sampling unit was the vehicle (e.g., car, truck, camper) that a person drove to a site. A two-stage sample was selected within strata. The first-stage units were clusters of vehicles (i.e., all vehicles available at a particular site on a given day). These were called site-days and were selected from a list by simple random sampling without replacement. The second stage sample consisted of the vehicles actually sampled for each site-day combination.

We identified eighty-five access points that met the criteria initially for the interviewing process at Lake Guntersville. Survey sites were designated recreational access points that were suitable for traffic counter equipment and safe for the interviewing teams. The sites were then stratified according to the main type of recreational activity conducted from each site. After surveying began, a few sites were eliminated from the sample as they were deemed unsafe or inaccessible due to seasonal weather conditions.

A total of 4,219 persons were contacted on-site, 4,021 of whom agreed to be interviewed, for a 95% response rate. Persons interviewed during the on-site survey were asked to provide their name and address for a follow-up CV mail survey. Of those interviewed, 2,715 persons (68%) provided the address information. Of these addresses, 155 were incomplete or incorrect and could not be used. This left a sample of 2,560 names and addresses for the CV survey.

Residential Survey

The residential survey was designed to collect lake use information from lake residents whose property was determined to have either direct or indirect legal access to the reservoir. In the questionnaire, lake residents were asked to provide detailed information on their recreational use occurring from their lake front or lake access residence.

An initial list of names and addresses, representing a complete inventory of lake property owners, was obtained from the TVA from an address database developed by the agency for a lake property owners study conducted just prior to ours. The database was adjusted by removing and adding those parcels that fit the necessary residential sampling criteria of either being located directly on the lake or with legal deeded access to the lake. When the database for the residential population was complete, the county tax assessors' offices for Marshall and Jackson Counties were contacted for mailing information, which was used to create a sample of

1,160 addresses. However, a total of 109 of these addresses were incomplete or incorrect and therefore not usable, reducing the final sample size for the residential survey to 1,051.

Contingent Valuation Survey

Names and addresses for the CV survey were obtained from the on-site and residential survey efforts described above. A total of 3,611 names and addresses were available (2,560 from the on-site survey and 1,051 from the residential survey).[1] The 3,611 CV questionnaires were mailed in the fall of 1992 following mail survey procedures discussed by Dillman. Initially, an explanatory cover letter, a CV questionnaire, and a postage-paid return envelope were sent to the respondents of the two earlier surveys. Approximately one week after the initial mailing, a post card reminder was mailed to all individuals in the sample. Two weeks after mailing the post cards, another cover letter, questionnaire, and return envelope were sent to those members of the sample who had not as of then responded. Of the 3,611 surveys mailed, 387 were returned as undeliverable, reducing the sample size to 3,224. The CV unit response rate was 50% (1,596) of this adjusted sample size.

A DC format was used to elicit WTP responses for plant management in the CV survey instrument. Five aquatic plant control alternatives were included in the questionnaire. These alternatives were developed in cooperation with the TVA and the USACE, co-managers of the lake. The first management scenario (MGTM) was a minimum control alternative under which aquatic plants would be allowed to grow naturally as if under ideal growing conditions. The coverage presented was 34,000 acres or approximately 50% coverage of lake surface area. All subsequent management alternatives were presented and valued relative to MGTM. Under Management Alternative A (MGTA), aquatic plant coverage was 20,200 acres (approximately 30% of lake surface area). Management Alternative B (MGTB) reduced the amount of aquatic plant coverage to 14,200 acres (approximately 20% of lake surface area). Management Alternative C (MGTC) set the level of plant coverage at 8,000 acres (approximately 12% of lake surface area). Finally, Management Alternative D (MGTD) set the amount of aquatic plants at near zero acres of surface area.

[1] We cross-referenced the on-site and residential survey address databases to eliminate duplication. Overlap between these two address databases was very small (e.g., less than ten cases).

Analysis of Survey Response-Related Bias

The variables used in the data analysis are defined in table 1. Income, age, family size, education, and an index of preference for plant coverage at Lake Guntersville are quantitative variables that were hypothesized to distinguish unit respondents from unit nonrespondents to the CV survey. A categorization of users as lake residents and nonresidents is a qualitative attribute hypothesized to be a distinguishing factor between respondents and nonrespondents.

Testing for Sample Nonresponse Bias

A simple test for sample nonresponse bias (SNB) on the unit level involves compar-

Table 1. Definition of Variables Used in Analysis

Variable	Description
USELAKE	Dependent variable; 1 if would use lake under management alternative, 0 otherwise
IRESPSC	Complete observation for scenario j; 1 if yes, 0 otherwise
URESP	Returned CV survey; 1 if yes, 0 otherwise
MGTA	Management alternative A; 1 if observation on A, 0 otherwise
MGTB	Management alternative B; 1 if observation on B, 0 otherwise
MGTC	Management alternative C; 1 if observation on C, 0 otherwise
MGTD	Management alternative D; 1 if observation on D, 0 otherwise
BID	Bid amount in questionnaire
INC	Midpoint of respondent's income range
EDUC	Index of respondent's education ranging from 1 (eighth grade or less) to 8 (doctoral degree)
AGE	Respondent's age
AGESQ	Respondent's age squared
FAMSZ	Number in respondent's household
PLNPREF	Index of respondent's preferences for aquatic plant coverage ranging from 1 (very small amount) to 5 (very large amount)
RES	Respondent a lake resident; 1 if yes, 0 otherwise

ing the distributions of characteristics of unit respondents and nonrespondents. If there are significant differences between the two, then the presence of SNB is indicated. However, unless there are data available on unit nonrespondents, the application of this test is limited to item respondents and nonrespondents. For the Lake Guntersville nonrespondents, these data are available from the pre-CV surveys, for both unit and item types. Thus,; we are able to estimate a unit response equation and test for differences between unit respondents and nonrespondents.

At this point, the objection may be raised: "What if the pre-CV survey is not representative of the population about which inference will be made?" If it is not, then the SNB problem may simply be pushed up to a higher level. However, where the survey design is appropriately stratified and the response rate is high, as in our case, we have strong reason to believe that the pre-CV survey is highly representative of the target population.

The results of the SNB tests are shown in tables 2 through 4. Table 2 applies to the overall sample without regard to user type. Tables 3 and 4 refer to fishers and nonfishers, respectively. In each case, significant differences exist between unit respondents and unit nonrespondents at all three levels of analysis. The implication is that the use of respondent statistics in evaluating the estimated WTP function would lead to SNB and therefore nonrepresentative estimates of welfare benefits.

Table 2. Tests for Sample Unit Nonresponse Bias, Overall Sample

	Respondents		Nonrespondents		
Variable	Mean	std	Mean	std	Test Statistic
INC	54,154	38,884	60,122	55,006	$t = 3.24$
AGE	50.4	14.6	42.5	15.4	$t = -13.42$
FAM$Z	2.7	1.2	2.6	1.3	$t = -1.97$
EDUC	4.4	1.7	3.7	1.6	$t = -10.95$
PLNPREF	2.9	0.9	2.9	1.3	$t = 0.99$
RES	0.20	–	0.06	–	$\chi^2 = 106.93$

Table 3. Tests for Sample Unit Nonresponse Bias, Fishers

	Respondents		Nonrespondents		
Variable	Mean	std	Mean	std	Test Statistic
INC	49,121	34,948	56,433	53,246	$t = 3.16$
AGE	50.1	14.4	43.6	14.6	$t = -8.31$
FAM$Z	2.8	1.2	2.6	1.2	$t = -2.67$
EDUC	4.2	1.6	3.4	1.5	$t = -8.40$
PLNPREF	3.2	0.9	2.6	1.2	$t = -9.78$
RES	0.16	–	0.01	–	$\chi^2 = 91.34$

Table 4. Tests for Sample Unit Nonresponse Bias, Nonfishers

	Respondents		Nonrespondents		
Variable	Mean	std	Mean	std	Test Statistic
INC	62,378	43,366	63,533	56,260	$t = 0.39$
AGE	50.8	15.0	41.6	15.9	$t = -10.31$
FAM$Z	2.7	1.2	2.7	1.3	$t = -0.07$
EDUC	4.8	1.7	3.9	1.6	$t = -8.95$
PLNPREF	2.5	1.0	3.6	1.2	$t = 14.14$
RES	0.24		0.11	–	$\chi^2 = 59.18$

Testing for Sample Selection Bias

The problem of sample selection bias (SSB) is addressed through model specification. The discussion of empirical results must therefore include both SSB and WTP. Before proceeding to the results of the estimation procedures, we briefly discuss specification of the WTP model. WTP models were specified and estimated for two distinct groups of users: fishers and nonfishers. It was hypothesized that fishers generally prefer more aquatic plant coverage because of their perception that aquatic plants improve fish habitat and angler success. On the other hand, nonfishers were hypothesized to prefer less aquatic plant coverage because of their concern that aquatic plants interfere with nonfishing activities such as water skiing and swimming.

It is often pointed out that qualitative response models in a random utility context should not be estimated without an intercept term unless there is "compelling reason to do so" (Greene). However, appealing to the economic theory involved in valuing changes in the level of a rationed, nonmarketed good, the total value curve should pass through the origin because it represents the initial combination of the good and income. We therefore omitted the intercept term from the WTP equation.[2]

The H2S procedure required the estimation of separate item selection (ITEM) and unit selection (UNIT) equations for the purpose of constructing variables to represent item and unit sources of SSB in the probit WTP equations. The probability of having a complete observation from a respondent on a scenario, i.e., item response, was specified as a function of AGE and EDUC, which were intended to capture familiarity with the lake and knowledge of its aquatic plant life. A quadratic in AGE was included to account for nonlinearities in the relationship between item-response and age. In addition, the management alternative dummy variables were included in ITEM to control for interdependencies among the observations. The UNIT response probability was assumed to depend on the opportunity cost of completing the survey, which we proxy with the variables EDUC and FAMSZ.

Table 5 shows the H2S results for fishers, while table 6 reports the results for nonfishers. Using difference-in-means t-tests, the management dummy coefficient estimates in ITEM for fishers are significantly different from each another. The same hypothesis test results held in ITEM for nonfishers. Given their significance and their difference from each other, they apparently captured, as intended, the effects of scenario specific heterogeneity in ITEM. EDUC, AGE, and AGESQ proved to be good predictors of response in ITEM for this data set, as did EDUC and FAMSZ in UNIT for nonfishers. However, FAMSZ did not aid in the prediction of fisher unit response.

Examination of the H2S primary equations, WTPI (WTP, Item) and WTPU (WTP, Unit), reveals that the coefficient estimates changed with the inclusion of the selectivity variables. This is an indication that the estimator of the WTP equation without selection (WTPNS) was biased. The greatest changes occurred in WTPU for both fishers and nonfishers. The estimated coefficients on the management dummy variables within each WTP equation were found to be significantly different from one another when testing the hypothesis with a difference-in-means t-test. Their significance and difference from one another may be an indication that respondents did perceive a difference among the various scenarios they were asked to value, and that there was scenario-level heterogeneity present in the data. Whether they indicate that there are significant differences among the values associated with each management scenario is discussed below. It must be kept in mind that these coefficients are intended to serve as indicators of whether the management scenario with which they are associated is valued differently from MGTM.

The coefficient estimates on the item and unit selectivity variables were significant for fishers and nonfishers alike, further indicating that the realized sample suffers from SSB. To interpret these estimates, it is important to realize that the estimates from the WTPNS models are based on a distribution truncated by the action of a selection rule. When that truncation is from below, as it is here, the truncated mean is biased in the direction of the coefficient on λ (Greene). The positive signs of the β_{λ}'s in the fisher selection models therefore indicate that the WTPNS model

[2] In general, including an intercept term does not lead to any meaningful differences in the parameter estimates. One exception is the fisher WTP equation, where including the intercept causes the coefficient estimates of management scenarios A and B to become insignificant. Parameter estimates for the intercept and no-intercept models are especially close for the regressions associated with unit non-response, which turns out to be the most empirically important source of non-response bias.

Table 5. Parameter Estimates, Heckman Two-Stage (H2S) Models, Fishers

Variable	ITEM	UNIT	WTPNS	WTPI	WTPU
MGTA	-0.0563	—	0.247	0.239	0.0920
	(-0.687)		(4.888)	(4.664)	(1.566)
MGTB	-0.162		0.152	0.137	-0.000320
	(-2.018)		(3.015)	(2.664)	(-0.005)
MGTC	-0.207	—	-0.0179	-0.0369	-0.172
	(-2.605)		(-0.353)	(-0.712)	(-2.924)
MGTD	-0.151	—	-0.261	-0.279	-0.407
	(-1.880)		(-5.172)	(-5.424)	(-6.922)
EDUC	0.0608	0.116	—		
	(4.128)	(7.395)			
AGE	0.0644	—		—	
	(17.061)				
AGESQ	-0.000766	—	—	—	—
	(-16.913)				
FAMSZ	—	-0.0551	—	—	—
		(-0.252)			
BID	—	—	-0.00262	-0.00271	-0.00184
			(-18.487)	(-18.821)	(-19.112)
INC	—		0.00442	0.00452	0.00228
			(8.993)	(9.045)	(6.290)
λ	—	—	—	0.137	0.258
				(2.407)	(5.064)
n	4,715	1,464	4,416	4,335	4,339
ln L	−1,529.17	-918.03	−2,822.59	−2,758.96	−2,758.96

Table 6. Parameter Estimates, Heckman Two-Stage (H2S) Models, Nonfishers

Variable	ITEM	UNIT	WTPNS	WTPI	WTPU
MGTA	-0.153	—	0.140	0.199	0.234
	(-1.462)		(2.100)	(2.872)	(3.002)
MGTB	-0.248		0.318	0.385	0.413
	(-2.413)		(4.665)	(5.415)	(5.211)
MGTC	-0.309	—	0.394	0.472	0.488
	(-3.037)		(5.713)	(6.501)	(6.126)
MGTD	-0.304		0.361	0.449	0.450
	(-2.993)		(5.317)	(6.273)	(5.708)
EDUC	0.0734	0.0941	—	—	—
	(4.239)	(16.890)			
AGE	0.0585	—		—	
	(12.108)				
AGESQ	-0.000687	—	—	—	—
	(-12.210)				
FAMSZ	—	0.0325	—	—	—
		(3.669)			
BID	—	—	-0.00341	-0.00343	-0.00261
			(-16.514)	(-16.346)	(-15.524)
INC			0.00401	0.00393	0.00363
			(7.700)	(7.472)	(7.817)
λ			—	-0.309	-0.153
				(-3.244)	(-2.343)
n	2,765	1,162	2,529	2,479	2,511
ln L	-983.84	-787.11	−1,556.29	−1,519.78	−1,543.69

overestimates the probability of a “yes” response to the DC question by not accounting for SSB. In turn, WTPNS overestimates WTP for fishers because fishers who hold lower values for decreasing plant coverage (e.g., prefer more plants) are under-represented in the realized sample. The negative signs of the β_λ's in the nonfisher models indicate that the WTPNS nonfisher model underestimates the probability of an affirmative DC response and WTP. Thus, WTPNS underestimates WTP for nonfishers because nonfishers who hold higher values for decreasing plant coverage (e.g., prefer less plants) are under-represented in the realized sample.

Estimation of the net economic benefits of aquatic plant management at Lake Guntersville, measured as mean willingness-to-pay (MWTP), closely followed the numeric integration procedure proposed by Hanemann. The estimates of MWTP were calculated for each user type. In estimating MWTP, variables were held at their means in the overall available sample—from all three surveys—not at the means of only those observations in the realized CV sample. This was done to correct for SNB. The variable *BID* was not held constant but was allowed to vary because it is the variable over which integration takes place. The limits of integration for the fisher and nonfisher models were found by solving the respective estimated probit WTP functions for the bid at which 80% of respondents would have been expected to answer “yes” to the DC question (lower limit) and the bid at which 20% of respondents would have been expected to answer “yes” (upper limit).

The results of the integration procedure are presented in figure 1 for fishers and figure 2 for nonfishers, for each model type and each management scenario. Given the signs of the estimated coefficients on the selection variables in the fisher H2S models, it would be expected that the estimated benefits would be greater for the model that does not account for sample selection, WTPNS, than for either of the two selection models, WTPI

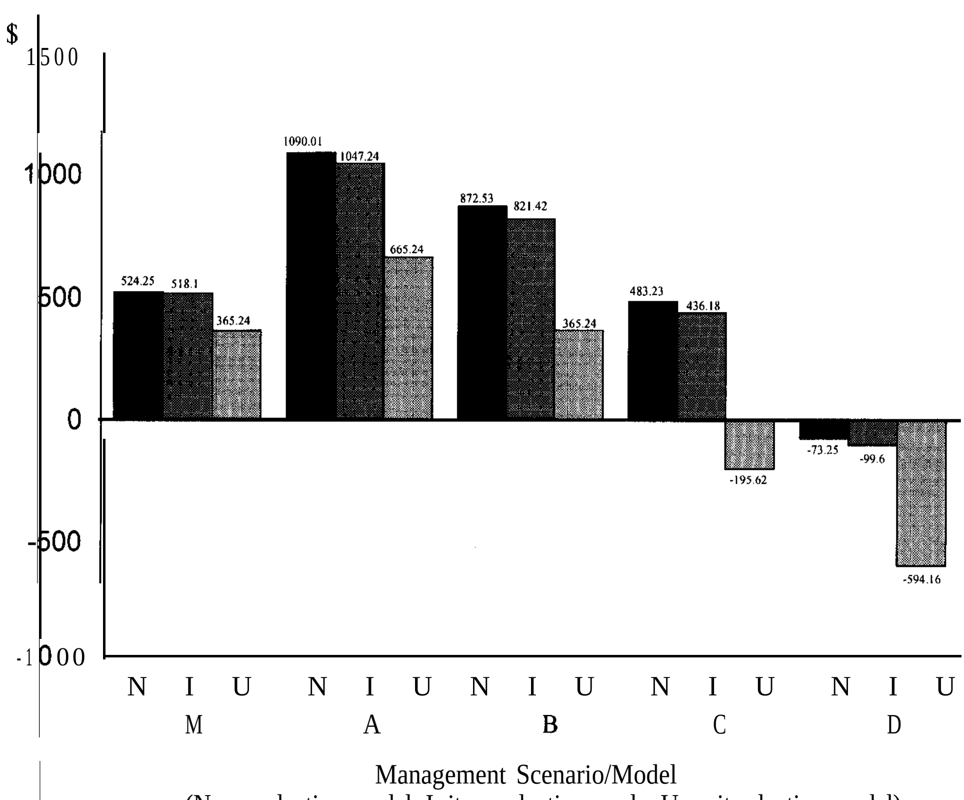

Figure 1. Mean willingness to pay estimates for H2S models, fishers

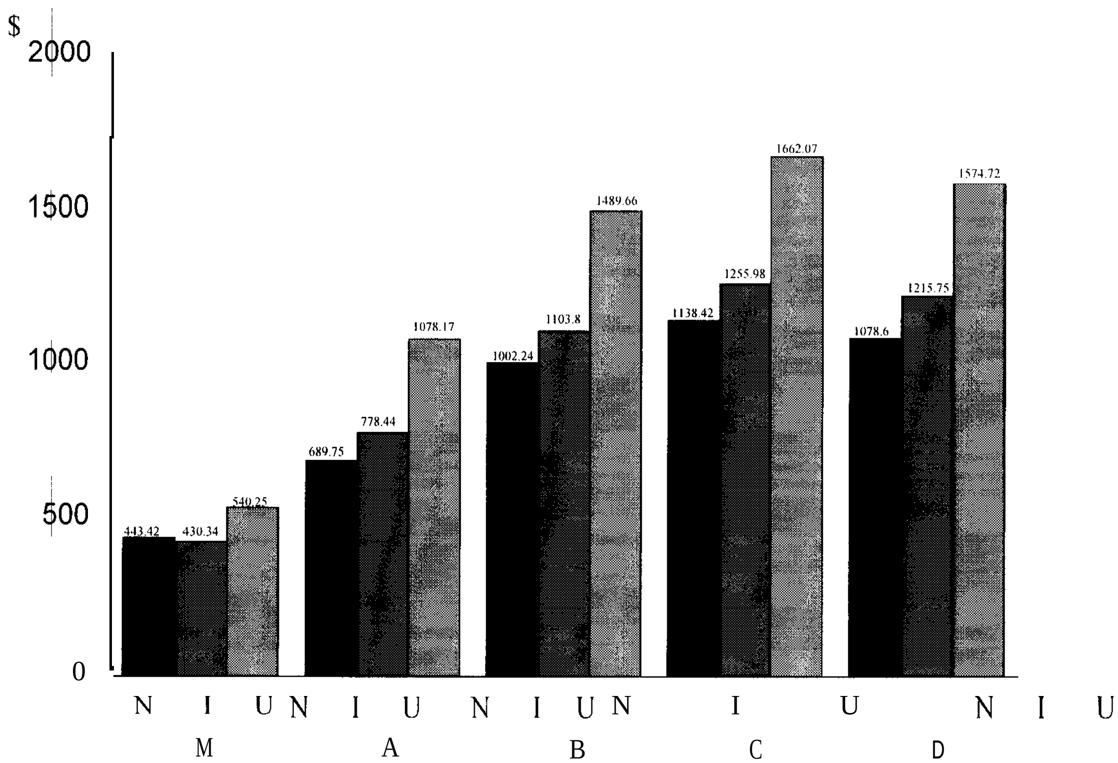

Figure 2. Mean willingness to pay estimates for H2S models, non&hers

and WTPU. Contrasting the no-selection estimates with the item-selection estimates, it can be seen that this expectation is borne out. The contrast is even more stark when considering the no-selection estimates relative to the unit selection estimates. The latter appear to indicate that fishers who hold lower values for decreasing the amount and distribution of the plants are greatly under-represented in the sample. Similarly, WTPNS for nonfishers underestimates MWTP relative to the selection models. The notable exception is MGTM between the no-selection and item-selection models.

Following a procedure described by Krinsky and Robb, 95% confidence intervals were calculated for the MWTP estimates (not shown). The confidence intervals suggest that the models without selection were seriously biased with respect to the unit source of SSB. For fishers, the confidence intervals for each scenario's WTPNS and WTPI models all included the other's point estimate. The only scenario for which the WTPNS and WTPU estimates may be said to be the same is MGTA. Results for the nonfisher models were similar. Overall, the indication is that, in these data, the bias introduced by item nonresponse does not seriously affect the estimation of welfare benefits. Unit nonresponse, however, had serious effects. This reinforces the earlier observation that the management scenario dummy variables within each model have coefficient estimates that are statistically different from each other. The significance of the selectivity coefficient estimates is also reinforced.

Implications and Conclusions

Economists are frequently asked to estimate the net economic value of a proposed or actual change in the level of an environmental amenity or natural resource. The contingent valuation method is often the technique of choice, but one criticism of the method is that the data used in estimation are obtained via survey methods. The validity of the method is questioned in large part because of the potential for survey response-related bias in data collection. Researcher awareness of the problem, what causes it, how to recognize it, and what means are available to mitigate its effects can improve the estimation of welfare benefits. In turn, the management and policy decisions based at least in part on such measures can be improved with less chance of misallocation of resources. Mitigation has a cost, however-obtaining data on the surveyed sample beyond that available from the CV survey.

The results reported in this paper suggest that tests for unit and item nonresponse bias should be routinely conducted in CV studies. When these types of biases are detected, the post-survey econometric solutions described in this paper appear capable of correcting for these biases. However, application of these solutions is dependent on what turns out to be missing in many CV applications-access to data describing nonrespondents. What can and should researchers do to improve access to such data?

Many CV applications, such as the one described in this paper, involve surveys of specialized populations where the researcher has to first conduct an inventory of the relevant population from which to draw a sample. Ideally, this inventory would capture and record information on 100% of the specialized population. As in the Lake Guntersville study, the inventory of a specialized population conducted by the researcher often becomes the source of names and addresses for follow-up surveys. To provide data necessary for applying the post-survey SRB techniques, researchers should collect information describing the basic characteristics of the specialized populations while conducting the initial inventory. This information would include socioeconomic variables and variables describing basic preferences. If this information is not collected as part of the initial inventory (e.g., if only names and addresses are collected), another option for collecting the data on unit nonrespondents is to conduct a follow-up telephone survey of nonrespondents.

CV studies often involve surveys of the general public. In these applications, a sample is drawn from a list of names and addresses thought to be representative of the general public. Such lists of names and addresses are often obtained from private, survey sampling firms. When dealing with a sample of the general public, a reasonably reliable source of data describing nonrespondents may be obtained from U.S. Bureau of the Census data. These data can then be applied with post-survey econometric techniques of the sort described in this paper to correct for nonresponse bias (e.g., see Cameron et al.).

When, for practical reasons, a researcher needs to conduct a partial inventory of a target population, there is always a danger that the partial inventory will not be representative of the population. For example, the inventory may itself suffer from unit nonresponse bias. Even inventory efforts targeted at an entire population such as the U.S. Census may suffer from nonresponse bias because of practical problems encountered in attempting to contact and interview all persons in a geographical area. Unfortunately, because in such cases we do not have information identifying and describing unit nonrespondents, unit nonresponse bias cannot be tested. The need for representative information on nonrespondents reinforces the importance of proper design and implementation of survey sampling techniques used to generate the inventory.

[Date submitted June 1999; accepted July 1999.]

References

Anderson, A.B., A. Basilevsky, and D.P.J. Hum. "Missing Data." ***Handbook of Survey Research,*** P.H. Rossi, J.D. Wright, and A.B. Anderson, eds. New York: Academic Press, 1983.

Cameron, T.A., W.D. Shaw, S.R. Ragland, S. Keefe, and J.M. Callaway. "A Method for Nonresponse Correction in the Analysis of Mail Survey Data." Unpublished manuscript, 1998.

David, M., R.J.A. Little, M.E. Samuhel, and R.K. Triest. "Alternative Methods for CPS Income Imputation." *J. Amer. Statist. Assoc.* 81(March 1986):29–41.

Davidson, R., and J.G. MacKinnon. ***Estimation and Inference in Econometrics.*** New York: Oxford University Press, 1993.

Dempster, A.P., N.M. Laird, and D.B. Rubin. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *J. Roy. Statist. Soc.* 39, no. 1(1977):1–38.

Dillman, D.A. ***Mail and Telephone Surveys.*** New York: John Wiley & Sons, 1978.

Edwards, S.F., and G.D. Anderson. "Overlooked Biases in Contingent Valuation Surveys: Some

Considerations." *Land. Econ.* 63(May 1987): 168-178.

Greene, W.H. *Econometric Analysis. New* York: Macmillan, 1993.

Hanemann, M.W. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *Amer. J. Agr. Econ.* 66(August 1984):332–41.

Heckman, J.J. "Sample Selection Bias as a Specification Error." *Econometrica* 47(January 1979):153–161.

Krinsky, I., and A.L. Robb. "On Approximating the Statistical Properties of Elasticities." *Rev. Econ. Statist.* 68(November 1986):715–19.

Loomis, J.B. "Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions." *Land Econ.* 63(November 1987):396–402.

Mitchell, R.C., and R.T. Carson. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington DC: Resources for the Future, 1989.

Orchard, T., and M.A. Woodbury. "A Missing Information Principle: Theory and Applications." *Sixth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1. Berkeley: University of California Press, 1972.

Whitehead., J.C. "Environmental Interest Group Behavior and Self-Selection Bias in Contingent Valuation Mail Surveys." *Growth and Change* 22(Winter 1991):10–21.

——, "Item Nonresponse in Contingent Valuation: Should CV Researchers Impute Values for Missing Independent Variables?" *J. Leisure Res.* 26(Third Quarter 1994):296–303.

Whitehead, J.C., PA. Groothuis, and G.C. Blomquist. "Testing for Non-response and Sample Selection Bias in Contingent Valuation." *Econ. Letters* 41(January 1993):215–220.

Whitehead, J.C., PA. Groothuis, T.J. Hoban, and W.B. Clifford. "Sample Bias in Contingent Valuation: A Comparison of the Correction Methods." *Leisure Sci.* 16(November 1994): 249-58.