A Note on Nonlinearity Bias and Dichotomous Choice CVM: Implications for Aggregate Benefits Estimation

R.A. Souter and J.M. Bowker

It is a generally known statistical fact that the mean of a nonlinear function of a set of random variables is not equivalent to the function evaluated at the means of the variables. However, in dichotomous choice contingent valuation studies a common practice is to calculate an overall mean (or median) by integrating over offer space (numerically or analytically) an estimated logit or probit function in which sample mean values for the concomitant variables are used. We demonstrate this procedure to be incorrect and we statistically test the procedure against the correct method for nonlinear models. Using data resulting in a well-behaved logit model, we reject the hypothesis of congruence between the two means. Such a finding should be considered in future single response dichotomous choice CVM studies, particularly when aggregation is of interest.

The contingent valuation method (CVM) is one of a battery of popular and accepted nonmarket valuation methods. In its various forms, the technique has been used in more than eleven hundred documented studies over the past twenty-five years to provide economic value information for nonmarket goods and services (Carson et al.).

The CVM procedure is based on eliciting individual willingness to pay (WTP) or willingness to accept (WTA) for a given change in the provision of a good or service. Depending on the wording of the elicitation method, one of the four Hicksian welfare measures is approximated (Mitchell and Carson). Typically, a valuation function for the average individual is estimated from a representative sample. For policy purposes, the welfare estimates are generally used (1) to estimate individual or group gains/losses within a given population or (2) to aggregate the gains/losses over all members of the population (Hanemann).

Since the late 1980s, the single response dichotomous choice (DC) or referendum method appears to be the most popular CVM procedure. Moreover, the recent NOAA Panel on Contingent Valuation concluded that the DC approach is generally preferred to the various elicitation alternatives (Arrow et al.).

Work by Bishop and Heberlein and by Hanemann is primarily responsible for triggering the adoption of DC by CVM practitioners. The DC procedure involves eliciting yes/no responses from individuals to randomly assigned monetary offers to accept or forego a given change in the provision of a good or service. Generally, parametric nonlinear statistical methods are applied to the yes/no data to model the probability of a yes (or just as easily, a no) response for a given offer amount and set of socioeconomic variables. The estimated probability function is then used to obtain median and mean economic surplus estimates. Mean WTP may be calculated analytically in the case of closed forms (Hanemann) or via numerical integration up to a truncation point (Duffield and Patterson; Ready and Hu). An alternative is to use the censored MLE approach described by Cameron and James. Interval estimates for either the conditional mean or median economic surplus may be obtained analytically (Cameron and James) or by numerical techniques such as the bootstrap (Duffield and Patterson) or Monte Carlo method (Krinsky and Robb).

While it has been argued that the choice of procedure for obtaining the mean with DC data may be mostly a matter of convenience (Park and Loomis), Duffield and Patterson provide a good

The authors are mathematical statistician and research social scientist with the USDA Forest Service, Southern Research Station, Forestry Sciences Lab, Athens, Georgia. Senior authorship is not assigned.

The authors acknowledge the helpful comments of the anonymous reviewers.

argument in favor of the truncated mean approach (TM). They contend that the TM is superior to the median and overall (analytical) mean on the grounds of (1) consistency with theoretical constraints, (2) statistical efficiency, and (3) ability to be aggregated. In the remainder the paper we use this approach without loss of generality.

Among the many other studies employing the TM are Sellar, Stoll, and Chavas; Boyle and Bishop; Bowker and Stoll; Boyle; Stevens, Glass, et al.; Stevens, Echeverria, et al.; Sun, Bergstrom, and Dorfman; Cordell and Bergstrom; Poe, Severence-Lossin, and Welsh; and Teasley, Bergstrom, and Cordell. The procedure followed by these researchers appears to correspond to Cameron, i.e., (1) estimation of the binary choice model (logit) on sample data, (2) collapsing the fitted probability model into two dimensions by creation of a grand intercept which is the sum of the model intercept and the products of the demographic variable parameter estimates and their associated sample means, and (3) integrating the area under the fitted two-dimensional cumulative probability curve up to an acceptable truncation point. The procedure described above is analogous to that used with linear models in which the overall mean may be correctly calculated at the means of the respective explanatory variables. However, in nonlinear models this miscalculation produces a bias (see appendix). When the estimate of interest, usually the mean or the median, is a nonlinear function of the explanatory variables, the correct approach is (1) to estimate the model(s) from the sample, (2) to integrate the fitted two-dimensional cumulative probability curves up to an acceptable truncation point for each individual, and (3) to take the average of estimated individual surplus estimates.

For aggregation over populations, e.g., a state, Loomis suggests substituting state average demographic variables for sample means. Such a practice retains the bias described above when used with nonlinear estimators. In the case of a population with a demographic distribution known to differ from a random sample, the nonlinear function must be integrated over an appropriate multivariate density.

Swallow et al. use a linear utility function to derive a nonlinear WTP estimate with parameters varying by demographic strata. They recognize the problem of substituting the population proportions of a state's demographic strata, which are state averages of indicator variables for demographic groups, into a nonlinear function to produce a state-level aggregate estimate. To ameliorate the bias caused by what they refer to as the "typical-preferences" approach, they appropriately recommend using a "typical-WTP" method that weights subpopulation estimates. They demonstrate up to a 7% difference by using an example based on weighting two subpopulation estimates.[1]

Below, we use a DC data set and a log-logistic functional form to demonstrate and statistically test for the incongruence of means calculated by what appears to be the common practice and the appropriate procedure for nonlinear models. We abstract away from the issue of heterogeneous preferences and model selection to focus attention on the basic issue of the calculation bias. Our findings suggest that the estimated means may be significantly different even in the case of an empirically well-behaved model. We limit our illustration to truncated means. The results also apply to analytically calculated integrals for dichotomous choice (e.g., Park and Loomis) and can be extended to the calculation of overall means for other commonly used nonlinear models such as the Tobit, which is often applied in open-ended CVM studies (e.g., see Reiling et al.).

Data and Methods

Data were obtained from a sample of on-site recreation users at the Lolo National Forest in Montana (USDA Forest Service). A dichotomous choice component was included as part of a larger visitor satisfaction questionnaire. A total of 202 users were interviewed over the summer of 1991.

The object of the CVM portion of the survey was to obtain information on annual individual net economic surplus associated with recreating at the Lolo National Forest. Interviewees were asked to consider their annual costs/expenditures for using this site. Next they were presented with a hypothetical situation in which their annual costs would have been increased by a given amount and were asked (yes/no) if they would still have used the site.[2] A follow-up question was asked of individuals answering no to identify possible protesters. Interestingly, no protesters were identified by this procedure; however, more than 25% of those sampled declined to complete the entire survey. Over 90% of the refusals were due to a decision not to answer the income question.

Our estimated logit model was specified as:

[1] It should be noted that in Swallow et al.'s illustration, both of the subpopulation estimates are biased because each is the result of linearly aggregating estimates.

[2] The increase in expenses structure is quite common to a number of published CVM studies. The survey questionnaire was extensively pretested and subjected to Office of Management and Budget approval. While there is always debate about survey questions, we think the data are acceptable for the purposes of our illustration.

Table 1. Maximum Likelihood Logit Parameter Estimates

Intercept	Ln *Offer*	*Income*	*Quality*	LRI[c]	Chi[2]	N
2.4407	−1.9041	.00003134	.7771	.4836	78.53	143
2.785[a]	.3662	.000017	.4404			
.3809[b]	.0001	.0583	.0777			

[a]Asymptotic standard errors.
[b]P-values associated with Wald chi-square.
[c]Likelihood ratio index (Greene, p. 682).

$$\text{Prob(yes)} = 1/(1 + e^{-(\alpha_0+\alpha_1 Income+\alpha_2 Quality+\beta \text{Ln} Offer)}) = [1 - F(x)] \quad (1)$$

where, e is the base of the natural logarithm, Ln *Offer* is the natural logarithm of the dollar amount for the dichotomous choice question, *Income* is individual annual gross income, and *Quality* is a Likert-type index of each individual's perception of overall recreation quality of the site; the α's and β are parameter estimates. $F(x)$ is the distribution function representing the probability of a no response to a given DC offer, x, where the offer is by definition greater than zero and the probability of a no response to a zero offer is zero. The results of the MLE estimation are reported in table 1.

The empirical model appears to be adequate for illustrating our point. The signs of the parameter estimates are consistent with theoretical expectations and are statistically significant. The likelihood ratio index indicates that this model fits the data as well as or better than most reported DC CVM studies.

Following Duffield and Patterson, the truncated mean WTP may be calculated as follows:

$$E(\text{WTP}_T) = \int_0^T (1 - F(x))dx = e^{-(\alpha_0+\alpha_1 Income+\alpha_2 Quality)/\beta} \Gamma\left(1 - \frac{1}{\beta}\right)\Gamma\left(1 + \frac{1}{\beta}\right) B\left(\frac{1}{T^{\beta}e^{(\alpha_0+\alpha_1 Income+\alpha_2 Quality)} + 1}, -\frac{1}{\beta}, 1 + \frac{1}{\beta}\right) \quad (2)$$

where symbols are as previously defined, and T is the upper value or truncation point of the distribution of allowable WTP, $\Gamma(g)$ is the gamma function evaluated at g, and $\mathbf{B}(l,p,q)$ is the probability that a beta-distributed random variable with parameters p and q is less than the limit, l—these probability functions are readily available in many mathematical programming languages.[3] By substituting in the mean income of $37,587 and the mean quality of 6.25, we employ this formula in the conventional sense, i.e., that of Cameron and of Cordell and Bergstrom, to obtain a truncated mean WTP of $131.71 (using a truncation point of $1,000). Alternatively, we obtain a truncated mean WTP of $142.59 when employing the correct approach. In this case, the difference in truncated means is 7.6%.

To examine the significance of this difference, we estimate 95% confidence intervals for the difference between the two methods using the Monte Carlo approach (table 2). Because this interval does not bracket zero, there is statistical justification to reject the hypothesis that the procedures are congruent. In the simulation, the average difference is 8.9%. Similar results were also obtained for median and nontruncated mean estimation from the two procedures.

Discussion

The practice of inserting explanatory variable means into nonlinear estimators to obtain an overall mean may be based on economists' long association with linear models in which the overall mean may be calculated at the means of the respective regressors. We illustrate the problem with a simple nonlinear function. We then empirically demonstrate that estimating a population's mean WTP by the common practice of evaluating the nonlinear function at the covariate sample means is incongruent with the correct procedure of averaging over the sample each individual's expected WTP. As well, we statistically test and reject at the $\alpha = .05$ significance level the hypothesis that the difference between the mean estimates is zero.

[3] This closed-form expression for the truncated mean is valid when the nontruncated integral is bounded, i.e., when the β parameter is less than -1; otherwise, a fat-tail problem occurs. When using bootstrapping or Monte Carlo methods for interval estimation, the formula is a useful indicator of the fat-tail problem.

Table 2. Truncated Means Difference Test

	Difference
Mean[a]	− 12.71
95% CI[a]	(− 29.5, − 1.34)

[a]Based on Monte Carlo simulation (1,000 replications).

In fact, for this particular Monte Carlo simulation, in which four maximum likelihood parameter estimates are jointly drawn from a multivariate normal density, each of the 1,000 parameter vectors produced a consistent sign of the bias between the two methods. This indicates that the linear analog of estimating the population mean WTP by evaluating the truncated mean integral at the sample means of the concomitant variables produces an underprediction bias arising from the inherent nonlinearity of individual WTP with the concomitant information. Of course, this bias is affected by the multivariate distribution of the concomitants in the population of interest—here, recreationists in the Lolo National Forest—and its magnitude will change from population to population.

In general these results can be extended to any CVM experiment where a nonlinear parametric procedure is involved, including DC medians and analytically calculated means as well as open-ended cases where a Tobit model is used. In our example the bias is on average just under 9%. This amount may or may not make the difference in a management decision; however, such a bias can be easily avoided by the appropriate application of aggregation methods. In addition, information provided by examining estimated WTP for each individual in the sample could alert the researcher to possible problems with the model and design space that would otherwise be overlooked. For example, unreasonable estimates may be readily identified for an individual with a certain set of characteristics.

Economic welfare analysis through the use of nonmarket valuation techniques is by no means an exact science that can be reduced to simple formulae. There may be situations where retaining a biased estimate of a population mean may be warranted for illustrative purposes. However, at a time when nonmarket methods are increasingly being used to guide public policy and, hence, are subject to more scrutiny, we think avoiding unnecessary bias by incurring a minor increase in computational expense would seem justified.

Appendix

Consider a simple case that deals with only two numbers, for example, $x_1 = 2$ and $x_2 = 3$, and a simple nonlinear function, $f(x) = 1/x$. At issue is whether the mean of $f(x)$ can be evaluated easily using the mean of x. Or more formally, does

$$\overline{f(x)} = f(\bar{x}),$$

where

$$f(x_i) = \frac{1}{x_i} \quad \text{and} \quad f(\bar{x}) = \frac{1}{\bar{x}},$$

with

$$\overline{f(x)} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \quad \text{and} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i?$$

Using the numeric values given, the answer is clearly no.

$$\overline{f(x)} \neq f(\bar{x}),$$

since

$$\overline{f(x)} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{3}\right) = \frac{5}{12}$$

and with

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{2}(2 + 3) = \frac{5}{2}$$

giving

$$f(\bar{x}) = \frac{1}{\bar{x}} = \frac{1}{\frac{5}{2}} = \frac{2}{5} \neq \frac{5}{12}.$$

In this simple case, the mean of a function is not equal to an estimate provided by evaluating the function at the mean of its arguments. This estimate is said to be biased. The bias, the difference between the actual mean of the function and its estimate, may be referenced as a percent of the actual mean. Here, we can determine percent bias, b:

$$b = \frac{\overline{f(x)} - f(\bar{x})}{\overline{f(x)}} * 100,$$

with

$$b = \frac{\frac{5}{12} - \frac{2}{5}}{\frac{5}{12}} * 100 = 4\%.$$

The bias is a 4% underprediction with this numerical example. This particular bias can be considered small or large, depending on the needs of the user and the particular application. However, this bias need not even be a question.

Considering the same framework of this problem, it is possible to establish the magnitude and direction of the percent bias given any two arbitrary numbers, say x_1 and x_2. It is easy to show that

$$b = \frac{r^2 - 2r + 1}{r^2 + 2r + 1} * 100,$$

where

$$r = \frac{x_1}{x_2}.$$

The percent bias is completely determined by the ratio of the two arbitrary numbers. Since this formula is always positive, the estimate will always be an underprediction if both numbers are positive, or an overprediction if both numbers are negative. If both numbers are of the same sign, then the percent bias is bounded to be no greater than 100% (e.g., if the numbers are 1 and 10, then the percent bias is 67%). If the two numbers are of differing signs and of similar magnitudes, then percent bias can be unbounded. In the trivial case, if the two numbers are equal, then the percent bias is zero.

Another consideration arises from a sampling framework. The aggregate estimate of a population mean that arises from two distinct strata, where the population is divided among the strata with given proportions, say p and $1 - p$, is often of interest. An attribute of interest may be nonlinearly related to other concomitantly measured attributes. The aggregate estimate of the population mean is the weighted mean over both strata. It is possible that the previously demonstrated nonequality of means might be offset by unequal proportions (namely, both p and $1 - p$ not equal to ½). It turns out that there are no proportions (weighted means) that can lead to equality of the population mean of the nonlinear function with the estimate provided by evaluating the function at the weighted mean of the concomitants for this particular problem's structure. This follows from a question similar to the question posed earlier. Are there weights, p and $1 - p$, such that

$$\overline{f(x)} = f(\bar{x}),$$

where

$$f(x_i) = \frac{1}{x_i} \quad \text{and} \quad f(\bar{x}) = \frac{1}{\bar{x}},$$

with

$$\overline{f(x)} = pf(x_1) + (1 - p)f(x_2) \quad \text{and}$$
$$\bar{x} = px_1 + (1 - p)x_2?$$

Using two arbitrary values, x_1 and x_2 (with ratio $r = x_2/x_1$), a value p is desired such that

$$\overline{f(x)} = f(\bar{x}).$$

So

$$\begin{aligned}\overline{f(x)} &= \frac{p}{x_1} + \frac{1-p}{r \cdot x_1} \\ &= \frac{1}{p \cdot x_1 + (1 - p) \cdot r \cdot x_1} \\ &= f(\bar{x}).\end{aligned}$$

This leads to

$$(r - 1)^2(1 - p)p = 0.$$

Again, if $r = 1$, then both strata are composed of the same values—the degenerate case—so any proportion, p, will satisfy the equation. Otherwise, the only proportion that would satisfy the equation is also from a degenerate case, namely, one stratum has a proportion of the population equal to zero. This demonstrates that, here, there is no population configuration that would allow an unbiased estimate of the population mean using a "plug-in" of the population mean of the function's arguments.

References

Arrow, K., R. Solow, P.R. Portnoy, E.E. Leamer, R. Radner, and H. Schuman. *Report of the NOAA Panel on Contingent Valuation.* Washington, D.C.: Resources for the Future, January 1993.

Bishop, R.C., and T.A. Heberlein. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(1979):926–30.

Bowker, J.M., and J.R. Stoll. "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." *American Journal of Agricultural Economics* 70(1988):372–81.

Boyle, K.J. "Dichotomous Choice, Contingent Valuation Questions: Functional Form Is Important." *Northeastern Journal of Agricultural and Resource Economics* 19(1990):125–31.

Boyle, K.J., and R.C. Bishop. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques." *American Journal of Agricultural Economics* 70(1988): 20–28.

Cameron, T.A. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15(1988): 355–79.

Cameron, T.A., and M. James. "Efficient Estimation Methods for 'Closed Ended' Contingent Valuation Surveys." *Review of Economics and Statistics* 69(1987):269–79.

Carson, R.T., N. Carson, A. Alberini, N. Flores, and J. Wright. "A Bibliography of Contingent Valuation Studies and Papers." La Jolla: Natural Resources Damage Assessment, 1993.

Cordell, H.K., and J.C. Bergstrom. "Comparison of Recreation Use Values among Alternative Reservoir Water-Level Management Scenarios." *Water Resources Research* 29(1993):247–58.

Duffield, J.W., and D.A. Patterson. "Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation." *Land Economics* 67(1991):225–39.

Greene, W.H. *Econometric Analysis*. New York: Macmillan, 1990.

Hanemann, W.M. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(1984):332–41.

______. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Reply." *American Journal of Agricultural Economics* 71(1989):1057–61.

Krinsky, I., and A.L. Robb. "On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics* 68(1986):715–19.

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

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

Park, T., and J. Loomis. "Comparing Models for Contingent Valuation Surveys: Statistical Efficiency and the Precision of Benefit Estimates." *Northeastern Journal of Agricultural and Resource Economics* 21(1992):170–76.

Poe, G.L., E.K. Severence-Lossin, and M.P. Welsh. "Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76(1994):904–15.

Ready, Richard C., and Dayuan Hu. "Statistical Approaches to the Fat Tail Problem for Dichotomous Choice Contingent Valuation." *Land Economics* 71(1995):491–99.

Reiling, S.D., K.J. Boyle, H. Cheng, and M.L. Phillips. "Contingent Valuation of a Public Program to Control Black Flies." *Northeastern Journal of Agricultural and Resource Economics* 18(1989):126–34.

Sellar, C., J.R. Stoll, and J.P. Chavas. "Validation of Empirical Measures of Welfare Change: A Comparison of Techniques." *Land Economics* 61(1985):156–75.

Stevens, T.H., R.J. Glass, T.A. More, and J. Echeverria. "Wildlife Recovery: Is Benefit-Cost Analysis Appropriate?" *Journal of Environmental Economics and Management* 33(1991):327–34.

Stevens, T.H., J. Echeverria, R.J. Glass, T. Hager, and T.A. More. "Measuring Existence Value for Wildlife: What Do CVM Estimates Really Show?" *Land Economics* 67(1991):390–400.

Sun, H., J.C. Bergstrom, and J.H. Dorfman. "Estimating the Benefits of Groundwater Contamination Control." *Southern Journal of Agricultural Economics* 24(1992):63–72.

Swallow, Stephen K., Thomas Weaver, James J. Opaluch, and Thomas S. Michelman. "Heterogeneous Preferences and Aggregation in Environmental Policy Analysis: A Landfill Siting Case." *American Journal of Agricultural Economics* 76(1994):431–43.

Teasley, R.J., J.C. Bergstrom, and H.K. Cordell. "Estimating Revenue-Capture Potential Associated with Public Area Recreation." *Journal of Agricultural and Resource Economics* 19(1994):89–101.

USDA Forest Service. "Customer . . . Understanding Today's Recreating Public (Lolo National Forest)." Athens, Ga.: Southern Forest Experiment Station, Outdoor Recreation and Wilderness Assessment Group, 1992.