RESOURCES
FOR THE FUTURE

Monte Carlo Benchmarks for Discrete Response Valuation Methods

Ju Chin Huang
V. Kerry Smith

Discussion Paper 97-23

February 1997

Resources for the Future
1616 P Street, NW
Washington, DC 20036
Telephone 202-328-5000
Fax 202-939-3460

Monte Carlo Benchmarks for Discrete Response Valuation Methods

Ju Chin Huang and V. Kerry Smith

Abstract

This paper argues that the widespread belief that discrete contingent valuation (CV) questions yield substantially larger estimates of the mean (and the median) willingness to pay (WTP) for nonmarket environmental resources in comparison to estimates from open-ended CV questions is unfounded. A set of Monte Carlo experiments estimate the factors influencing the performance of WTP estimates based on discrete response models. Most of the error in the WTP estimates arises from the specification errors that are common in most of the empirical models used in the literature. These experiments suggest models based on choices where WTP is dominated by non use (or passive use) values are likely to have smaller errors than where large use values influence these decisions.

Key words: discrete response contingent valuation, Monte Carlo, non-market valuation

JEL Classification Nos.: C93, D12, Q2

Table of Contents

List of Tables

Monte Carlo Benchmarks for Discrete Response Valuation Methods

Ju Chin Huang and V. Kerry Smith*

I. INTRODUCTION

An unsubstantiated judgment about the performance of discrete response questions for contingent valuation has received widespread acceptance in the literature. This conclusion holds that (closed ended) contingent valuation (CV) questions yield substantially larger estimates of the mean (and the median) willingness to pay (WTP) in comparison to estimates based on open-ended or (payment card) responses. These upward "biases" are more pronounced, it has been argued, when the WTP is dominated by passive use value (i.e. nonuse value).[1] It has also been suggested that problems may arise when individuals do not have choice experience. While this is a widely accepted view of the properties of valuation estimates based on discrete response models, it is hard to isolate the precise source of these conclusions.

At least three types of modeling decisions influence the performance of discrete response models in estimating the WTP for environmental resources. Two of these arise in implementing the economic model required to use a censored response for estimating an unobserved continuous random variable (i.e. the WTP). First, for parametric models, the analyst must select a specific function to characterize people's budget constrained preferences. It is this model that describes how an individual's choices relate to the object of choice and terms presented in each decision. There are additionally modeling decisions that we have considered part of the model specification but could easily be treated separate. They relate to the issues to be resolved in implementing a model for discrete response data. They include specifying the economic and non-economic determinants assumed to influence a respondent's

---

* Assistant Professor, Department of Economics, East Carolina University and Arts and Sciences Professor, Duke University and Resources for the Future University Fellow. Partial support for Smith's research was provided the UNC Sea Grant Program R/MRD-32.

[1] This is one implication suggested to explain the Kealy-Turner [1993] and Cummings et al. [forthcoming] results.

choice. For the economic factors, the theory of consumer behavior provides reasonable guidance. However, for economic variables outside the conventional (simple) description of choice there is little to provide a basis for these judgments. The situation is even worse for the non-economic attitudinal and information variables.

The second aspect of the decisions made in using discrete response models stems from the analyst's specification for the stochastic error used to characterize discrepancies between the model's predictions and the observed outcomes. Decisions about the error structure are usually made to simplify estimation. At best, they should be interpreted as providing a basis for diagnostic analysis of models, rather than the theoretical basis for valuation estimates.

Finally, the format of the questionnaire, text of the information provided, and the design of the survey (or experiment) establish a context that specifies the circumstances of each choice, the object of choice, as well as the procedure for reporting choices for the individuals composing the sample.[2] The relative importance of these three components of the economic modeling process in determining the properties of the resulting WTP estimates is not clear. Nonetheless, the literature has tended to assume that discrepancies in measures of the WTP for specific resources across studies are due to the survey-related components of the process (i.e., the third decision in our taxonomy) and not to the first two modeling decisions.

The purpose of this paper is to suggest that this conclusion is unwarranted. Resolution of the persistent questions about the properties of WTP estimates derived from discrete response models requires that the sources of error in discrete response methods to be isolated and "benchmarked" against a known standard. Of course, we should acknowledge that this goal is easier to offer as a standard than to implement in practice. Analyses based on people's choices, whether CV or revealed preference, will never have access to the true WTP. Even demand revealing mechanisms used in experimental studies require maintained assumptions.[3] Among the important assumptions is that the participants understand the incentives

[2] See Kopp and Smith [forthcoming] for a summary of how these basic concepts in the micro economic theory underlying individual choice relate to the design of CV questions.

[3] See Harrison [1996] for a thorough discussion of these assumptions.

underlying the proposed choice mechanisms. Multiple practice rounds with other choices along with some type of "training" has sometimes been required for the more complex incentive systems. We don't understand how this "experience" influences behavior. It appears that it does yield differences in the experimental results.[4]

To make progress in understanding the differences in valuation estimates between methods that rely on continuos versus censored responses, we must first have a standard to evaluate the size of the error in WTP estimates expected with discrete response methods. As we noted, this standard has been absent from all discussions of the problems with this approach. It can be estimated in a straightforward way. Indeed, the logic of what we propose could be implemented as part of the evaluation of results from any discrete response CV study based on actual data. They are developed assuming a "true" specification for preferences as well as the constraints to choice, and then simulating choices under different terms. These choices are then used to recover estimates of WTP which can be compared to the "true" WTP for those objects of choice. This paper reports such an evaluation, describing the results of a Monte Carlo study that and offers the discrete CV counterpart to the controlled evaluations of travel cost (Kling [1988], Adamowicz et al. [1989]) and hedonic property value (Cropper, Deck, and McConnell [1988]) models in the literature.

Our results suggest that the size of the error in estimating WTP from discrete response models (measured as the root mean squared error for estimates of WTP relative to the "true" value) is influenced by the factors directly associated with the incomplete nature of the specification of discrete choice models, the size of passive use to use values, and the structure of true preferences. These factors can cause substantial variation in the proportionate error in probit or logit estimates of WTP for a change in quality. Differences in people's income and access conditions to the use-related resources account for a large fraction of the proportionate errors in comparison to variations in the parameter choices that characterize consumer preferences and control the relative size of passive use to use values.

[4] See Cummings et al. [1995] for discussion of one set of experiments.

After a brief review of the evidence on the performance of discrete response CV estimates of WTP versus open ended surveys in the next section, we outline the design for our experiments in Section III. Section IV summarizes the criteria used to evaluate probit and logit estimates of WTP based on the censored responses characteristic of discrete choice CV questions. The last section discusses the implications of our findings for general conclusions about discrete versus open-ended CV question models.

II. DISCRETE VERSUS OPEN-ENDED CV RESPONSE

The existing evidence on the properties of discrete CV questions is mixed. The record is a mixture of applications and experiments, with limited experimental control and quite different practices being treated as either discrete response methods or open-ended questions. Because of this tendency to define rather loosely membership in each approach to eliciting valuation responses, each additional comparison seems to add to the flux in the research record without offering decisive evidence on the properties of estimates from the discrete CV format.[5] Another reason for these outcomes is the incomplete control over all the modeling decisions that can influence WTP estimates in both applications and experiments (whether laboratory or simulated market studies).

Table 1 summarizes a selection of the results of eight studies comparing discrete choice (DC) and open ended (OE) contingent valuation questions. Four studies stand out as yielding the largest values for the ratio of mean WTP estimated with open ended to that from discrete choice questions. These are Sellar et al. [1985], McFadden [1994], Holmes and Kramer [1995], and Ready et al. [1996]. Of these four, the McFadden work is the most detailed. The ratios reported in Table 1 understate the disparities he found. We report comparisons of the medians because his results seem to have been greatly influenced by the

[5] An example of this tendency to adopt simplified and sometimes confusing classifications is the designation of WTP values elicited with a payment card as an example of the open-ended format. Those from a conjoint study comparing two situations have been treated as discrete choice studies. While there are clearly elements in the format that would support both judgments, it is not clear that we know enough about how all the features of the question format influence whether these judgments are neutral to a comparative evaluation of how each variation in question mode influences the discrete response versus open ended questions.

treatment of the tails of the mixed log normal distribution used to estimate his parametric models.[6] The ratios of DC to OE estimates of mean WTP ranged from about 23 to 73; suggesting that the adjustments used in practice (i.e. truncating the distributions and deleting outliners) may have reduced these ratios. There are also several important contrasting features between his analysis and these other studies. The McFadden analysis used the same preference specification to analyze both the OE and DC response models for the utility function. This is not the case for the Sellar et al. and Ready et al. analyses. As a result, differences in findings across question modes in these other studies reflect both the different maintained models hypothesized to describe respondent's preferences and the type of question used. It is not possible from the information reported in them to separate the effects of each of these potential choices.[7] This criticism also applies to the Johnson et al. [1990] study as well as to a number of others not included in the table.

Several of the remaining studies have different problems further narrowing the set of evidence that can be used to support the common beliefs about upward bias in discrete choice CV questions. For example, Kealy and Turner [1993] and Kristrom [1993] ask both OE and DC questions to the same respondents.[8] While Kealy and Turner account for this potential correlation in their statistical model analyzing their data, both the order effect and the error correlation (between the responses to the two questions) were significant factors in some of the models for the public and private goods. This would suggest that the mean computations need to consider the effects of order and the more generally dependent nature of the samples for their conclusions about OE and DC estimates. This adjustment was not done.

---

[6] All of McFadden’s discrete response models yield mean WTP estimates with exceptionally wide confidence intervals, suggesting the roles for income and the proposed payment in the models do not offer plausible explanations for respondent’s choices.

[7] The Ready et al. model does not estimate WTP. Because the questions relate to a single grapefruit of different types and ask for the maximum price per unit (for the payment card) or state a price per unit for the discrete choice, the quantity response must be jointly modeled to estimated WTP (see Eom and Smith [1994]). To the extent there are differences in these quantity responses across question models these disparities would confound further their comparison.

[8] Kristrom asked a portion of his sample both OE and DC questions and a separate component only OE questions. His analysis appears to pool responses on OE questions from both groups.

Holmes and Kramer's study seems to raise a different problem. A large proportion of estimated difference in WTP, using the models based on the actual data seems to be due to the larger error variance with the discrete choice response model. This is not a new issue. It is a reason for long-standing concerns about the sample size and associated cost of surveys with discrete CV questions. Their split samples have comparable sizes (192 for DC and 186 for the payment card). Because models for both modes assumed a log normal distribution, estimates of the mean WTP will be biased downward (due to Jensen's inequality). While they apply Goldberger's [1968] proposed correction term, this adjustment reduces but does not eliminate the bias.

The ratio of corrected $W\hat{T}P_{DC}$ to $W\hat{T}P_{OE}$ is 9.4 while the ration of the ln ($W\hat{T}P_{DC}$) to ln ($W\hat{T}P_{OE}$) (both unbiased estimates) is only 1.71, suggesting that the two approaches do a more comparable job in gauging percent changes in the WTP with changes in the relevant determinants. An important source of this difference in the estimated levels for WTP appears to be in the greater variability in the DC model's error in comparison to the open-ended model's error. The Goldberger correction is

$$WTP_i = exp(X_i\hat{\beta}_i) \cdot exp(\hat{\sigma}_i^2 / 2)$$

We can use the separation of estimates of location and scale effects in the correction to consider the size of the error variance in comparison to the economic component of the model. The ratio of $(exp(\hat{\sigma}_{DC}^2 / 2))/(exp(\hat{\sigma}_{OE}^2 / 2))$ is 3.17, implying that the deterministic component's contribution is about 2.98. While this remains large, it is not as extreme as their corrected means would imply. Holmes and Kramer's simulations also suggest larger WTP estimates with the discrete choice model applied to the case with a true set of parameters set by the estimates from the open-ended sample. What seems to have been overlooked is the relatively close correspondence between estimates of the effects of each economic determinant between the two actual samples. This is also true for comparisons between the DC model using the simulated data and the parameters serving as true values from the OE estimates with the actual data. All of this seems to suggest that the large discrepancy in this case arises from a composite of the log transformation together with the known larger error variance associated with models based on censored or DC data.

Thus, for a variety of reasons we are left with two studies providing the primary basis for -- DC and OE response comparisons -- McFadden and Boyle et al. [1996]. McFadden finds order of magnitude differences and Boyle et al. find no pronounced difference in the ratios and only one significant difference in the means.[9] McFadden's analysis also evaluated differences in the amount of public good offered, considering four different specifications of the wilderness areas opened for logging. The DC models do not appear to describe respondent behavior. None of the models (across quantities of wilderness preserved) displays a significant effect of income entered in a form that restricts the effects of income and the proposed cost of the plan to be equal. Boyle et al. do find significant effects for the proposed cost and do not report tests for the effects of income. Neither study explores the role of conventional factors we might expect to influence choices. For example, in McFadden's analysis one might have expected consideration of direct measures of access, such as travel cost (rather than using the state of residence as a proxy for whether respondents would have been users of the areas that were proposed for logging). The same type of issue would seem to have been a plausible variable for inclusion in the model used with the moose hunting sample considered by Boyle et al.

It is difficult to resolve the discrepancy in the findings of these two studies. McFadden's analysis casts doubt on all CV approaches, but most especially those using discrete choice questions. Boyle et al.'s models do not support the types of differences between open-ended and discrete choice question modes found in McFadden's work.[10] Both studies evaluate the models with split samples and consistent models. Thus, despite the widespread acceptance of a bias in DC estimates, past studies offer only one comparison (McFadden) with strong evidence. We do not know from what is reported in this study whether these discrepancies are due to the other respondent and use-related factors

---

[9] It should be noted that the skewed nature of the open-ended data and the variability in the discrete choice models were important reasons for their conclusions.

[10] This contrast seems even more dramatic when it is recognized that the oil spill component of the Boyle et al. work was developed as part of the <u>same</u> larger scale evaluation of CV as the McFadden's analysis, see Hausman [1992].

influencing CV choices or to the generic character of the question that was used in this study. Indeed, the overall lack of support from the DC models for any economic model could be used as support for a conclusion that the study does not necessarily question DC but rather the ability of respondents to deal in this format with the choices being posed. Thus, we conclude from this literature, as well as other studies we investigated but do not summarize, that there appears to be clear motivation for estimating a benchmark for their approximate influence of modeling decisions on DC estimates.

III. MONTE CARLO METHODS FOR BENCHMARKING DISCRETE RESPONSE MODELS

A. Background

Our argument that simple evaluations of DC versus OE estimates are incomplete is based on the premise that the estimates derived from choice data reflect multiple specification errors (even if the process we assume describes how people respond to CV questions corresponds exactly to economic choices). This judgment follows from the fact that simple (usually linear) models are frequently used to analyze the data from DC studies. The models assume other relevant variables, such as measure of access to the resources that may be crucial to the choice can be ignored (or assigned to a limited role). Similarly, the role of income is often ignored in the simple models by assuming a locally constant marginal utility of income.

To investigate the importance of these approximations we specify three different models each with preference specifications that allow for use and passive use values for a quasi-fixed (or rationed) good that is assumed to represent an environmental resource. With given income and relative prices we solve each model for the WTP function describing how each hypothetical (or simulated) individual's monetary valuation of the quasi-fixed good changes with income and relative prices. These simulated values are altered for different specifications of the parameters of the preference functions to allow the relative size of use and passive use value for the rationed good to change across experiments.

These models are used to construct choices describing how each "individual" would respond to discrete CV questions. This response process assumes stochastic error is included in evaluating the "value" of the change in the rationed good with the proposed "fee" in comparison to the base conditions. This evaluation underlies the constructed WTP function. With these simulated choices, the fees, and the characteristics of these "individuals," we estimate linear choice models, comparable to what has been done in the literature, and evaluate the implied WTP in comparison to the true values.

B. Preference Specifications

The three behavioral models describe different types of consumption patterns. All three models assume that each "constructed" individual spends his (or her) entire income on two commodities, the environmental quality-related good x_1 and a composite good x_2 that represents the consumption of all other goods. The level of the quasi-fixed good which we treat as representing the environmental quality consumed by the individuals when consuming x_1 is represented by R. R is assumed to be an essential good. That is, unless R exceeds a threshold value R_m, the consumption of x_1 generates no satisfaction to the individual. For example, beach trips do not provide individuals any enjoyment unless beach quality is above certain level. To simplify the analysis, R is assumed to be greater than R_m in all experiments. It is assumed that environmental quality enters the models through its influence on the parameters in each preference specification. The three utility functions used in our experiments are given in equations (1) through (3):

$$U_{i1} = a_1 x_{1i} + ln(x_{2i}) + b \tag{1}$$

$$U_{i2} = a_1 x_{1i} x_{2i} + x_{2i} + b \tag{2}$$

$$U_{i3} = a_1 x_{1i} + a_2 x_{2i} + a_3 x_{1i}^2 + a_4 x_{1i} x_{2i} + a_5 x_{2i}^2 + b \tag{3}$$

We assume that: $a_1 = \alpha_1 (R-R_m)$, $a_3 = \alpha_2 (R-R_m)$, and $b = exp(\text{b}(R-R_m))$ and thus allow the quality-related good to interact with the market goods through the parameters to each function. All these parameters are constant for a given level of R. However, when R is proposed to change, the parameters (a_1, a_3, and b) change and the nature of the choices of the market goods change.

Each numerical specification of α_1 controls the degree of the linkage between x_1 and environmental quality *(R)*. Thus, α_1 also impacts the size of the use value for changes in *R*. The parameter, β, controls the magnitude of the passive use value for changes in environmental quality. Each specification for preferences treats the contribution of *R* associated with passive use as separable from marketed goods. This formulation is consistent with Hanemann's [1988] definition and implies increments to *R* will increase utility provided that both β and $R\text{-}R_m$ are greater than zero. These effects are realized even though x_1 may be zero. By controlling the relative magnitude of α_1 and β, it is possible to represent an array of different environmental resources with differing mixes of use and passive use values. The budget constraint for each model is: $I_i = p_{1i}\, x_1 + x_{2i}$, where *I* and p_1 are normalized by the price of x_2. The indirect utility function for each model can be derived from solving the constrained utility maximization problem.[11] The effect of nonuse [$b = exp(\beta\, (R\text{-}R_m\,))$] remains (strongly) separable in the indirect utility functions for each model.

The first utility function, presented in equation (1), is additive in x_1 and x_2. The environmental resource is also separable from x_2. Madariaga and McConnell [1987] used this model to illustrate the measuring of passive use (or existence) and use values for an essential resource. The total willingness to pay (labeled here as $WTP_{\mathrm{i}1}$) for an increase in environmental quality from R_0 to R_1 can be derived by setting the two realized utilities under each set of conditions equal and solving for the one-time payment consistent with this indifference. This is given in equation (4).

$$WTP_{i1} = \frac{(R_1 - R_0)I_i}{R_1 - R_m} + \frac{p_{1i}}{\alpha_1(R_1 - R_m)}\left[\ln\left(\frac{R_0 - R_m}{R_1 - R_m}\right) + e^{\beta(R_1 - R_m)} - e^{\beta(R_0 - R_m)}\right] \tag{4}$$

WTP in this form includes both use and passive use values.

The second utility function, given in equation (2), has a cross product term linking x_1 and x_2. This cross product term restricts x_1 to be a weak complement with x_2. This implies

---

[11] A detailed derivation is available on request from the first author.

that x_1, the environmental quality-related commodity, has no value if consumption of the composite good, x_2, is zero. Thus, in this case, use of the environmental resource is no longer separated from that of the other goods. By equating two alternative conditions of the utility function characterizing what the different quality levels imply for the choice, the true WTP for consumer *i* (labeled WPT_{i2} for this function) can be derived and is given in equation (5) below.

$$WTP_{i2} = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$\begin{aligned}
\textit{where} \quad & A = \alpha_1^3 (R_0 - R_m)(R_1 - R_m)^2 \\
& B = -2\alpha_1^2 (R_1 - R_m)^2 I_i - 2\alpha_1 (R_1 - R_m) p_{1i} \\
& C = \alpha_1 (R_0 - R_m)(\alpha_1 (R_1 - R_m) I_i + p_{1i})^2 - \alpha_1 (R_1 - R_m)(\alpha_1 (R_0 - R_m) I_i + p_{1i})^2 \\
& \qquad + 4\alpha_1^2 (R_0 - R_m)(R_1 - R_m) p_{1i} (e^{\beta(R_1 - R_m)} - e^{\beta(R_0 - R_m)})
\end{aligned} \tag{5}$$

The other root of the WTP function would imply values for WTP that are greater than I_i, and is therefore not an economically feasible solution.

The third model, presented in (3), is a quadratic utility function that has been used in several empirical studies (e.g., Cameron [1992], Kealy and Bishop [1986]). The parameters a_1, a_3, and b are assumed to be quality related. The other parameters are assumed constant and the condition $a_3 a_5 - a_4 > 0$ is imposed, so the utility function is concave. In this model the consumption of x_1 and x_2 is not complementary. Use of the environmental resource is assumed to make a separable contribution to preferences from that made by x_2, as in model 1. The expression for the total WTP (labeled as WTP_{i3}) is more complex. It is given in equation (6) below.

$$WTP_{i3}=\frac{E_1+2F_1I_i+\sqrt{(E_1+2F_1I_i)^2-4F_1^2\left[D_1-D_0+e^{\beta(R_1-R_m)}-e^{\beta(R_0-R_m)}+(E_1-E_0)I_i+(F_1-F_0)I_i^2\right]}}{2F_1^2}$$

$$\begin{aligned}
&\textit{where } D_0=D(R_0);E_0=E(R_0);F_0=F(R_0) \qquad (6)\\
&D_1=D(R_1);\ E_1=E(R_1);\ F_0=F(R_1)\\
&D=D(R)=\frac{-(a_2p_{1i}-\alpha_1(R-R_m))^2}{2G}\\
&E=E(R)=\frac{\alpha_1(R-R_m)(2a_5p_{1i}-a_4)+a_2(2\alpha_2(R-R_m)-a_4p_{1i})}{G}\\
&F=F(R)=\frac{4\alpha_2(R-R_m)a_5-a_4^2}{2G}\\
&G=G(R)=2a_5p_{1i}^2-2a_4p_{1i}+2\alpha_2(R-R_m)
\end{aligned}$$

The negative root is economically irrelevant. The complexity of model specification has increased from model 1 to model 3. This has lead to a corresponding complexity in the expressions for the WTP for the postulated change in environmental quality. The transformations to the additive errors, introduced to the indirect utility functions in modeling "individuals'" choices are progressively more complex with each model. These expressions, in turn, influence the implied distribution for the WTP based on each preference specification. All three structures satisfy the conventional assumptions of preferences in specific ranges for parameter values. The selection of parameter values in our sampling experiments are based on existing empirical studies so the plausibility of preference structures can be ensured.

The impact of b on the welfare measure is positive in all three models, implying those who have a higher passive use value for the increases in *R, ceteris paribus,* are willing to pay more for that quality improvement. Passive use value does not influence the demands for goods, the marginal utility of income, or the price and income elasticities of demand for the private good. The parameter, a_1, is the channel in each model for controlling the importance of the use value associated with changes in *R*. In all three models, a_1 directly affects consumption of both goods. The relative impact of the size of a_1 on WTP depends on resource levels, R_1 and R_0, and the importance of passive use value, b.

C. Experimental Design

The experiments were designed to control the relative size of passive use to use values while altering the linkages between nonmarket and the market goods through the three preference specifications. The set of household incomes and relative prices for the market goods is fixed for all three preference specifications and fixed across the repeated samples that compose each Monte Carlo experiment. Each sample consists of 200 observations. The values for Cameron's [1992] parameter estimates were used to set the parameters for the quadratic model with the parameter vector (a_2, a_2, a_4, a_5) specified to be: *(-.0013674, 3.309, .002579, -.2334)*. The parameters in the other models are derived from the specifications of a_1 and b.

The choice process is assumed to follow Hanemann's [1984] utility difference model. Choices are based on differences in the indirect utility function with the change in *R* (*i.e.,* R_1- R_0) and a specified fee. These are treated as fully describing the circumstances of choice that are conveyed to each person. Because an additive error is included with 1 these utility differences, the induced WTP distributions for each preference specification are heteroscedastic. The initial errors for choices are assumed to be in standard normal form. In each experiment, 100 independent drawings (replications) of errors with 200 values in each sample are generated. Each experiment consists of a preference specification (i.e., one of the three preference models given in equations (1) through (3)), a parameter setting, and the fixed set of values for income and relative prices. Choices require a set of "fees" for the quality improvements. These are randomly selected from a set of twenty values *(.001, .05, .10, .15,..., .95)* that are assigned to each pair of income and relative price and are fixed in repeated samples. The values for income were drawn from a normal distribution with mean five and standard deviation of one. The values for the relative prices are drawn from an independent uniform distribution *(.1, .6)*. Twelve sets of parameter values are selected varying a_1 over six values *(.03, .04, .05, .06, .07, .08)* and b over two values *(.02, .10)*. The values for the non-market environmental resource used in defining WTP, (R_0, R_1) are fixed at *(20, 20.5)* with $R_m=10$.

Table 2 constructs for each parametric specification the average ratio of passive use to use values implied across the values of the independent variables and the parameter values used for each model. The numbers in parentheses below each average ratio is the standard deviation in the ratio. These ratios will vary with the values for income and relative prices across the 200 observations (intended to represent constructed "individuals"). Increasing b by five times increases the relative size of passive use value to use value by about ten times. With model 3 we see there is much greater variability in these ratios across observations comprising the sample.

D. Implementation

In practice, the analyst does not know the appropriate specification for the choice model. One of the most common approaches relies on a linear WTP function for a specified improvement in quality from R_0 *to* R_1 as in equation (7).[12]

$$WTP_i^* = \gamma_1 + \gamma_2 I_i + u_i \tag{7}$$

where u_i is a random error. WTP_i^* is not observed. As a consequence this specification maintains that a respondent's choice provides a bound for WTP_i^*. If an individual is asked to pay t for improving the environmental quality from R_0 to R_1 and responded "yes", then WTP^* is at least t. A "no" answer implies that t is an upper bound for WTP^*. This is the conventional description of the discrete choice framework and implies that a probit or logit specified in terms of I_i and t will identify sufficient parameters to estimate $E(WTP_i^*)$ as in (8)

$$E(WTP_i^*) = (\hat{\gamma}_1 + \hat{\gamma}_2 I_i) / \hat{\gamma}_3, \tag{8}$$

where the $\hat{\gamma}$'s are the estimated coefficients from the logit or probit model.[13]

We apply linear models to each sample from each of the twelve parameterizations of our three models. Two estimators -- probit and logit -- are used with these models. Because each of the behavioral models provides a different value of the true WTP_i for each respondent, we have 2,400 separate true values for WTP and associated evaluations (i.e., 12

---

[12] The model was first proposed by Cameron [1988]. McConnell [1990] demonstrated that for the case of a linear indirect utility function, it is equivalent to Hanemann's [1984]model.

[13] g_1 is the intercept, g_2 the coefficient of I_i, and g_3 the coefficient of t, that is usually interpreted as an estimate of the reciprocal scale parameter for u_i.

parameterizations x 200 observations in one sample) for each model (and estimator). One hundred replications for the independent draws of the random errors are generated in each parameterization of the models.

To facilitate our evaluation of these results, we use a summary measure for the error in the WTP estimates to combine results across replications. It provides a gauge of the error due to the use of a discrete choice model (as an approximation). It is the root mean square error of WTP across replications for each of the 200 "constructed individuals". It is measured relative to the true WTP for each "individual." By expressing the root mean square error of the estimated WTP (est WTP_{ik}) as a proportion of the true WTP (true WTP_i), denoted PRTMSE, the deviation is normalized for comparison across "constructed individuals." The expression for this measure is given in equation (9):

$$PRTMSE_i = \frac{\sqrt{\frac{1}{m}\sum_{k=1}^{m}(estWTP_{ik} - trueWTP_i)^2}}{trueWTP_i} \quad i = 1, \ldots, n, \tag{9}$$

where m is the number of replicates in the sampling experiments and n is the sample size.

To summarize the results we estimate response surface equations with the performance index (PRTMSE) specified as a function of the features distinguishing each "constructed individual," the design point used as the fee, the estimator (probit or logit) and the model parameterization. This approach allows the results to indicate the contribution each feature of the experiment makes to the proportionate error in the estimates using discrete choice models for measuring WTP.

The regression analysis of PRTMSE pools findings across estimators for each model and across models. As a result, there are 4,800 observations for each model (i.e., 2400 x 2 estimators). In the case of model 3, some observations were deleted as technically feasible but economically implausible because they implied negative use values.[14]

IV. RESULTS

Our summary of the performance of DC models recognizes that sampling experiments are being used to evaluate the properties of probit and logit used with specification errors (i.e. assuming linear choice functions when the true form is both non-linear and includes more variables). This evaluation is based on the properties of the implied estimates for WTP which is a non-linear function of the estimated parameters. Of course, because both probit and logit and maximum likelihood (ML) methods for their assumed error distribution, their implied estimates of the WTP are also ML estimates. These properties can be expected to differ with preference specification as well as with the parameterization used in generating each experiment. Each influences the relative size of the specification errors involved with the estimating model in relationship to the models generating the true WTP that underlie the observed choices. This impact is most easily recognized for the first utility specification. As indicated in equation (4), WTP is linear in income (I) and relative prices (p_1) for a given incremental change in the environment quality. Thus, in this case the use of a linear choice function in terms of the proposed fee and income is nearly consistent with underlying preferences. If the model had included p_1 it would have conformed with the theoretically correct specification for a choice function with this experiment.

Therefore it should not be surprising that the proportionate error increases as the importance of the omitted term increases. This is seen in Table 3 with the first two summary

[14] The negative use values associated with model 3 occur in all six a_1 values. The distribution is given as follows:

α_1	Number of Negative Use Values
.03	88
.04	58
.05	39
.06	22
.07	14
.08	10

These arise for both settings for b and with both estimators. As a result we deleted 924 (231 x 4) observations associated with these negative use values.

equations. p_1 has a positive and significant effect on the size of the error in estimates of WTP (based on the linear estimating model used to derive the expression in equation (8)) relative to the true value. The level of income tends to decrease the error and the size of the fee posed increases it. The decision to use probit or logit is unimportant regardless of model selected to generate the data.

One of the most interesting aspects of these results is found in the term reflecting how the setting of parameter values for each model control the size of use and passive use values in total WTP. This is indicated through the relative size of b to a_1 and has a negative and significant effect on the proportionate error in estimates of WTP. Moreover, if we use the preference function for each model to calculate the implied true ratio of passive use to use values (a term likely to be correlated by construction with income, p_1 and potentially the denominator of the dependent variable), we see the same effect for Models 1 and 3. That is, as passive use value increases related to use value the error in estimates of WTP from linear models with only income and the fee decline. This conclusion is interesting because it stands in sharp contrast to current judgments about the conditions that affect the properties of discrete choice CV.

The explanation of this contradiction between our findings and the guidance in current literature is direct. Our regression models omit a term related to use value (i.e. p_1). The larger passive use is to use, the less important use related factors are to choices and, in turn for the estimates of WTP. It is important to note that this effect does not address the motivation underlying concerns about DC with larger passive use. For the most part, in these cases analysts are concerned about how experience and knowledge of the environmental resources affects respondents' ability to deal with CV type questions. For environmental resources that do not lead to significant use values, it may well be the case that the CV survey provides the first time a respondent has heard about that specific resource and the proposed change to it. Under these circumstances will these individuals' responses describe how they would decide with an actual choice where they may have had more "advance warning" and knowledge? Our design requires choices to come from the specified preferences, and thus there is not a basis for observing incomplete consideration or understanding of the proposed change in these simulated choices. Thus, we are not offering information on this issue. Instead, the findings are suggesting that this type of behavior is not the only source for error in discrete response models with incomplete specification of the use-related determinants, as was the case in both the McFadden and Boyle et al. studies. For our results the importance of the use-related variables becomes a gauge of the specification errors associated with ignoring them. Thus, as passive use grows in size relative to use values, these potential specification errors decline.

Model 2 changes matters because the link of weak complementarity between x_1 and x_2 spreads the effects of *R*. Little intuition can be derived from the expression for the true WTP. Income continues to reduce proportionate errors and p_1 to increase them, but it is only significant when the parameters (b/a_1) are used to characterize passive use to use effects. When the relative values are included, p_1 is no longer influential alone for the proportionate errors in WTP. In this case the larger the passive use to use, the larger the errors. However, this is not for the reasons this popular judgment envisions. Rather, in this model the connections imposed to other goods (e.g. x_2) changes the nature of the specification error associated with our simple estimating model.

Finally, as we implied above, with model 3 the use related linkage between x_1 and x_2 is eliminated, the pattern reverts to that of model 1. Larger values for passive use to use values reduce proportionate errors. All the other effects except the proposed fee remain comparable.

Pooling across preference specifications enhances the significance of the estimated coefficients but very likely misrepresents the relative effects of individual properties of the data and of preferences. Indeed, the pooled model would be rejected because of pronounced differences in the individual parameters in comparison to those for the response surfaces with the individual preference specifications. We present it to illustrate that some consistent summary judgments (in terms of the signs of major influences to estimates of WTP) would be possible even in situations where the summaries pooled across different preferences either due to differences in the resources involved or unobserved heterogeneously in the people.

V. IMPLICATIONS

There are at least two important implications of these experiments. The first involves a confirmation of the questions we posed about drawing general conclusions concerning the properties of DC models based on the results in the literature. Conventional theory suggests that a number of factors should influence people's choice functions. The models estimated with discrete choice CV responses traditionally have been quite simple, under the assumption that the omitted terms are either unimportant and do not vary over the sample. Our findings indicate even fairly small model specification changes can influence the size of the proportionate error in estimates of WTP. The models in Table 3 provide a simple benchmark for the range of errors that can easily span most of the reported differences between OE and DC results. Moreover, the OE estimates are themselves random variables and a full evaluation would require consideration of these errors as well. To develop such an evaluation we would need to incorporate some framework to describing peoples' responses to open-ended questions within a behaviorally consistent model for WTP. One possibility, comparable to the one we have used, would be to assume they arise as selections from a payment card. Following Cameron and Huppert [1991] we would expect the errors in selections to be based on the coarseness in the grid of values of the card. Under this scheme it seems reasonable to expect a comparable set of influences due to omitted variables. Because these approaches have better ability to describe the WTP distribution, we would expect smaller estimates for the error variances in the fitted models. This does not necessarily mean they are less sensitive to the specification biases due to omitting factors important to respondents' valuations for environmental quality.

In the context of our experiments with discrete choice models that acknowledged the links between use values and the importance of specification errors that can arise with simple estimating models, passive use values tend to reduce the errors, <u>not</u> increase them. Of course, in this case large passive use values imply the variables generally linked to uses are less important to the modeling of choice function and in turn to the properties of WTP estimates.

As a practical matter we do not know either the preference structure or relative importance of passive use values in each application, and this brings us to the second implication of our experiments. Incomplete information about preferences and about the process people use to answer discrete CV questions does not preclude constructing a counterfactorial experiment. This counterfactorial can be used to benchmark what to expect with DC estimates. Instead of using the synthetic data format as in Boyle et al. and Holmes and Kramer, this would entail generating Monte Carlo data from a preference specification consistent with the estimating model as well as the parameter estimates derived from each specific application and its DC data. Based on these sampling experiments (with variations in key model parameters over a range of values that are consistent with each application), a response surface in terms of the proportionate errors in the implied WTP measures could be estimated. Such a function would provide some basis for gauging one component of the errors introduced by the censoring inherent in DC responses. It also allows the effects of model misspecification encountered in CV applications to be evaluated on a case by case basis.

TABLE 1. A Sample of Discrete-Choice and Open Ended Comparisons

Authors	DC/OE WTP	Results Based on Test	Model Assumption	Error Assumption	Commodity
Boyle et al. [1996]	1.2 - 1.4 0.4 - 1.9[a]	yes significant difference for one of four samples	open ended and DC asked of different respondents	parametric	moose hunting prevention of oil spills
Ready et al. [1996]	3.6 - 4.4 (DC versus payment card)	yes	open ended and DC asked of different respondents	parametric for model	grapefruit with different risk due to pesticides
Holmes and Kramer [1995]	9.4 (DC/payment card)	yes significant difference	payment card and DC asked of different respondents	normal	protect spruce-fir forests in Southern Appalachian
McFadden [1994]	DC>OE (order of magnitude) 5.0 - 10.6 (Medians)	yes significant difference	open ended and DC asked of different respondents	nonparametric and parametric	preserving wilderness
Kealy and Turner [1993]	1.03 - 1.1 1.4 - 2.6	yes significant difference	both questions asked of each respondent	parametric (normal)	<u>Candy bar</u> reduction in acid rain damage in Adirondacks
Kristrom [1993]	DC>OE mean not computed	Test of difference in distributions	a subset of sample received both OE and DC	nonparametric	preserving forest in Sweden
Johnson et al. [1990]	1.6	no	open ended and DC asked of different respondents	normal and logit	access to whitewater river recreation
Sellar et al. [1985]	5.7 2.6	Tests of difference in Means reported for all methods	open ended and DC asked of different respondents	logit, log-normal	access fee Livington take Houston take

[a] Range for the oil spill sample reflects uncensored and censored sample distinctions

TABLE 2. Relative Magnitude of Passive-use and Use Values[a]

Model 1						
α_1=	.03	.04	.05	.06	.07	.08
β=.02	0.085 (0.067)	0.057 (0.038)	0.043 (0.027)	0.034 (0.021)	0.029 (0.017)	0.024 (0.014)
β=.10	0.967 (0.761)	0.643 (0.434)	0.483 (0.305)	0.387 (0.235)	0.324 (0.191)	0.278 (0.161)
Model 2						
α_1=	.03	.04	.05	.06	.07	.08
β=.02	0.060 (0.059)	0.042 (0.037)	0.033 (0.027)	0.027 (0.022)	0.023 (0.019)	0.020 (0.016)
β=.10	0.685 (0.669)	0.480 (0.417)	0.374 (0.311)	0.308 (0.251)	0.262 (0.211)	0.228 (0.182)
Model 3						
α_1=	.03	.04	.05	.06	.07	.08
β=.02	0.739 (1.049)	0.902 (3.684)	0.551 (2.284)	0.382 (2.045)	0.144 (0.318)	0.091 (0.145)
β=.10	8.396 (11.920)	10.245 (41.837)	6.258 (25.933)	4.334 (23.224)	1.638 (3.605)	1.033 (1.643)

[a] The elements in the table are the average values of the ratio of passive use to use values for each model and parameter setting. The values depend on the income and relative price values. The numbers in parentheses are the standard deviations.

TABLE 3. Response Surface Models for PRTMSE[a]

	Model1 A	Model1 B	Model2 A	Model2 B	Model3 A	Model3 B	Pooled A	Pooled B
Intercept	0.8816*** (54.7862)	0.8785*** (55.3531)	4.1699*** (39.6336)	4.3204*** (39.4703)	5.6237*** (23.6601)	4.8265*** (19.7938)	2.4329*** (29.2267)	2.2526*** (27.6227)
Income	-0.0968*** (-37.7762)	-0.1062*** (-40.3570)	-0.5816*** (-34.7093)	-0.5482*** (-29.8810)	-0.7061*** (-17.1057)	-0.7743*** (-17.9837)	-0.4300*** (-32.4786)	-0.4337*** (-32.6356)
p_1	0.0238 (1.2365)	0.1323*** (6.4334)	0.2645** (2.1037)	0.0611 (0.4551)	2.2333*** (7.4926)	2.8843*** (9.2024)	0.6782*** (6.8978)	0.7284*** (7.3588)
fee	0.0232** (2.4660)	0.0227** (2.4265)	0.0631 (1.0250)	0.0631 (0.0091)	0.1811 (1.3323)	0.1781 (1.2553)	0.0908* (1.9110)	0.0900* (1.8874)
Probit Dummy	0.00286 (0.5107)	0.00286 (0.5134)	0.0066 (0.1805)	0.0066 (0.1776)	0.0190 (0.2337)	0.0190 (0.2240)	0.0088 (0.3121)	0.0088 (0.3110)
Passive-use/use		-0.1094*** (-14.4080)		0.3016*** (4.8884)		-0.0093*** (-3.4689)		-0.0060*** (-3.6151)
b/a_1	-0.0364*** (-12.4579)		0.2573*** (13.4696)		-0.8567*** (-18.8330)		-0.1510*** (-10.0493)	
Model2 Dummy							1.3221*** (39.4080)	1.3217*** (39.2678)
Model3 Dummy							1.4611*** (40.8306)	1.4885*** (41.1945)
R^2	0.2501	0.2580	0.2255	0.2002	0.1586	0.0843	0.1940	0.1888
F Statistic	319.856	333.392	279.125	239.929	145.8543	71.2480	463.149	447.6816
d.f.	4794	4794	4794	4794	3870	3870	13468	13468

[a]Numbers in parentheses are the ratio of the estimated coefficient to the estimated standard error for testing the null hypothesis of no association.
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

REFERENCES

Adamowicz, W. L., J. L. Fletcher, and T. Graham-Tomasi. 1989. "Functional Form and the Statistical Properties of Welfare Measures," *American Journal of Agricultural Economics* 72 (May), pp. 414-421.

Alberini, Anna. 1995. "Testing Willingness-to-Pay Models of Discrete Choice Contingent Valuation Survey Data," *Land Economics*, vol. 71, no. 1, pp. 83-95.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extra-market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics*, 61, pp. 926-930.

Boyle, Kevin J., F. Reed Johnson, David W. McCollum, William H. Desvousges, Richard W. Dunford, and Sara P. Hudson. 1996. "Valuing Public Goods: Discrete Versus Continuous Contingent Valuation Responses," *Land Economics*, 72 (August), pp. 381-396.

Cameron, Trudy A. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation of Censored Logistic Regression," *Journal of Environmental Economics and Management*, 15 (Sept.), pp. 355-379.

Cameron, Trudy A. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Non-market Goods," *Land Economics*, 68 (August), pp. 302-317.

Cameron, Trudy A. and David D. Huppert. 1991. "Referendum Contingent Valuation estimates: Sensitivity to the Assignment of Offered Values," *Journal of the American Statistical Association*, vol. 86, no. 416, pp. 910-918.

Cooper, Joseph C. and John Loomis. 1992. "Sensitivity to Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models," *Land Economics*, 68 (May), pp. 211-24.

Cooper, Joseph C. and John Loomis. 1993. "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models: Reply," *Land Economics*, 69 (May), pp. 211-24.

Cropper, Maureen L., Leland Deck, and Kenneth E. McConnell. 1988. "On the Choice of Functional Form for Hedonic Price Functions," *Review of Economics and Statistics*, 70 (November), pp. 668-675.

Cummings, Ronald G., Steven Elliott, Glenn W. Harrison, and James Murphy. 1997. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* (forthcoming).

Cummings, Ronald G., Glenn W. Harrison, and Laura L. Osborne. 1995. "Can the Bias of Contingent Valuation be Reduced?: Evidence from the Laboratory," Economics Working Paper B-95-03, University of South Carolina.

Eom, Y. M. and V. Kerry Smith. 1994. "Calibrated Nonmarket Valuation," unpublished paper under revision, Center for Environmental and Resource Economics, Duke University.

Goldberger, Arthur S. 1968. "The Interpretation and Estimation of Cobb Douglas Functions," *Econometrica*, 36 (July-October), pp. 464-472

Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66, pp. 332-341.

Hanemann, W. Michael. 1988. "Three Approaches to Defining 'Existence' or 'Non-Use' Value under Certainty," working paper, Dept. of Agricultural & Resource Economics, University of California, Berkeley, July.

Harrison, W. Glenn. 1996. "Experimental Economics and Contingent Valuation," Discussion Paper B 96-10, Department of Economics, University of South Carolina, October.

Hausman, Jerry A., ed. 1992. *Contingent Valuation: A Critical Assessment* (Amsterdam, North Holland).

Holmes, Thomas P. and Randall A. Kramer. 1995. "An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous Choice Contingent Valuation," *Journal of Environmental Economics and Management*, 29 (July), pp. 121-132.

Johnson, Rebecca. L., N. Stewart Bregenzer, and Bo Shelby. 1990. "Contingent Valuation Question Formats: Dichotomous Choice Versus Open-Ended Responses," in R. L. Johnson and G. V. Johnson, eds., *Economic Valuation of Natural Resources: Issues, Theory, and Applications* (Boulder, Colo., Westview Press).

Kealy, Mary Jo and Richard C. Bishop. 1986. "Theoretical and Empirical Specification Issue in Travel Cost Demand Studies," *American Journal of Agricultural Economics*, 69 (August), pp. 660-667.

Kealy, Mary Jo and Robert W. Turner. 1993. "A Test of the Equality of Close-Ended and Open-Ended Contingent Valuation," *American Journal of Agricultural Economics*, 75 (May), pp. 321-331.

Kling, Catherine L. 1988. "Comparing Welfare Estimates of Environmental Quality Changes from Recreation Demand Models," *Journal of Environmental Economics and Management*, 15 (September), pp. 331-340.

Kopp, Raymond J. and V. Kerry Smith. 1996. "Constructing Measures of Economic Value," in R. J. Kopp, W. Pommerhrne, and N. Schwarz, eds., *Determining the Value of Non-Market Goods: Economic Psychological and Policy Relevant Aspects of Contingent Valuation* (Boston, Mass., Kluwer Nishoff).

Kristrom, Bengt. 1993. "Comparing Continuous and Discrete Contingent Valuation Questions," *Environmental and Resource Economics*, 3 (February), pp. 63-72.

Madariaga, Bruce, and Kenneth E. McConnell. 1987. "Exploring Existence Value," *Water Resources Research*, vol. 23, no. 5, pp. 936-942.

McConnell, Kenneth E. 1990. "Models for Referendum Data: The Structure of Discrete Models for Contingent Valuation," *Journal of Environmental Economics and Management*, 18, pp. 19-34.

McFadden, Daniel. 1994. "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, vol. 76, no. 4, pp. 689-708.

Ready, Richard C., Jean C. Buzby, and Dayuan Hu. 1996. "Differences Between Continuous and Discrete Contingent Value Estimates," *Land Economics*, 72 (August), pp. 397-411.

Sellar, Christine, John R. Stoll and Jean-Paul Chavas. 1985. "Valuation of Empirical Measures of Welfare Change: A Comparison of Non-Market Techniques," *Land Economics*, 61 (May), pp. 156-175.