



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Agricultural & Applied Economics Association

Contingent Valuation and Social Choice

Author(s): Daniel McFadden

Source: *American Journal of Agricultural Economics*, Vol. 76, No. 4 (Nov., 1994), pp. 689-708

Published by: [Oxford University Press](#) on behalf of the [Agricultural & Applied Economics Association](#)

Stable URL: <http://www.jstor.org/stable/1243732>

Accessed: 22/10/2011 18:38

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Agricultural & Applied Economics Association and *Oxford University Press* are collaborating with JSTOR to digitize, preserve and extend access to *American Journal of Agricultural Economics*.

<http://www.jstor.org>

Contingent Valuation and Social Choice

Daniel McFadden

The contingent valuation method for estimating the existence value of natural resources is examined for psychophysical robustness, statistical reliability, and economic sensibility. Extensions of standard models for willingness-to-pay, and suitable econometric techniques for analyzing these models, are developed. The analysis is applied to a series of experiments on the value of preserving wilderness areas in the western United States. The results call into question the reliability of the CV method for estimating existence values.

Key words: contingent valuation, willingness-to-pay, social choice, existence value.

How can you measure the net benefits to society from actions that impact environmental resources? An economist's answer is to employ Hicksian consumer surplus, determining the equivalent variation in income that leaves each consumer indifferent to the action. When consumers are rational and consumer surplus can be measured reliably from market demand functions, this is a satisfactory basis for welfare calculation, subject to the customary caveats about distributional equity and consistency if compensation is not actually paid.

When externalities, public goods, or informational asymmetries interfere with the determination of consumer surplus from market demand functions, one can try to set up a hypothetical market to elicit an individual's equivalent variation, or willingness-to-pay (WTP). This is called the contingent valuation method

(CVM). The approach elicits stated preferences from a sample of consumers using either open-ended questions that ask directly for WTP, or referendum (closed-ended) questions that present a bid or a sequence of bids to the consumer, and ask for a yes or no vote on whether each bid exceeds the subject's WTP. A single referendum experiment presents only one bid; a double referendum experiment presents a second bid that is conditioned on the subject's response to the first bid, lower if the first response is no and higher if it is yes.

An extensive literature has investigated the use of CVM to value environmental goods, and in recent years has promoted it for evaluation of goods such as endangered species and wilderness areas whose value comes primarily from existence rather than active use.¹ The typical CVM experiment in environmental economics asks about a single commodity, often with a fairly abbreviated or stylized description that assumes the consumer can draw upon prior knowledge. Typically, there is no training of the consumer to reduce inconsistent (e.g., intransitive) responses, or to reconcile responses with the consumer's budget. In recent years, referendum procedures have tended to replace open-ended elicitations, apparently because they circumvent a relatively high incidence of

Daniel McFadden is a professor in the Department of Economics, Econometrics Lab, University of California, Berkeley.

This research was supported in part by Exxon Company, U.S.A., as part of its program of legal defense in environmental cases. The initial design of the experimental study described in this paper was done by Peter Diamond, Jerry Hausman, and the author. The detailed design and implementation of the study was done by Elizabeth Hoffmann of the University of Arizona and Mike Denning of Exxon Company. The empirical analysis in this paper extends work carried out jointly with Greg Leonard, and reported in McFadden and Leonard (1993). I have benefited from discussions with Mike Denning, Bill Desvousges, Trudy Cameron, Peter Diamond, Jon Goldstein, Jerry Hausman, Danny Kahneman, and Paul Ruud, and from the comments of referees. Research assistance was provided by Pinghua Young and Li Gan. The opinions expressed in this paper are those of the author, and not necessarily those of the Exxon Company or any of the individuals named above.

Review coordinated by Richard Adams.

¹ See, for example, Bishop and Heberlein; Bowker and Stoll; Cameron and Huppert; Cameron and James; Carson, Hanemann, and Mitchell; Carson; Cummings, Brookshire, and Schultz; Hanemann; Hanemann, Loomis, and Kanninen; Loomis (1990); McConnell; Duffield and Patterson; Park, Loomis, and Creel; McClelland et al.

nonresponse or "protest" response found in open-ended studies.

In assessing CVM, there are three common-sense questions that can be asked: (a) Is the method psychometrically robust, in that results cannot be altered substantively by changes in survey format, questionnaire design, and instructions that should be inconsequential when behavior is driven by maximization of rational preferences? (b) Is the method statistically reliable, in that the distribution of WTP can be estimated with acceptable precision using practical sample sizes? Reliability is a particular issue if CV surveys produce extreme responses with some probability, perhaps due to strategic misrepresentation. (c) Is the method economically sensible, in that the individual preferences measured by CVM are consistent with the logical requirements of rationality (e.g., transitivity), and at least broadly consistent with sensible features of economic preferences (e.g., plausible budget shares and income elasticities)? CVM might fail to meet these criteria because respondents receive incomplete information on the consequences of the available choices, or are given inadequate incentives to be truthful and avoid strategic misrepresentation, or because the experimental design is not sufficiently rich to detect and compensate for systematic and random response errors. Beyond such technical problems, there could be a fundamental failure of CVM if consumers do not have stable, classical preferences for the class of commodities, so that the foundations of Hicksian welfare analysis break down. Intuitively, the further removed a class of commodities from market goods where the consumer has the experience of repeated choices and the discipline of market forces, the greater the possibility of both technical and fundamental failures. The broad sweep of evidence from market research, cognitive psychology, and experimental economics suggests that the existence value of natural resources, involving very complex commodities that are far outside consumers' market experience, will be vulnerable to these failures (McFadden 1986).

The following sections discuss, in turn, a series of statistical issues in analyzing WTP data, parametric methods for estimating mean WTP, an experiment that was designed to detect and quantify technical failures of CVM, and the results from the experiment.

Statistical Issues in CV Data Analysis

This section considers the following statistical questions: What are the merits of parametric versus nonparametric estimates of mean WTP? What are the advantages of conditioning on covariates rather than analyzing marginal WTP data? What is the impact of outliers, and how should they be handled?

Parametric Versus Nonparametric Analysis

An elementary estimator of population mean WTP can be obtained without parametric assumptions by taking a random sample of the population, eliciting a truthful WTP from each respondent using an open-ended question, and forming the sample mean of these stated values. Provided the population variance of WTP is finite, the variance of the mean WTP estimator is inversely proportional to sample size.

Using referendum questions complicates matters only slightly, since votes at a sufficiently broad and closely spaced range of bid levels can be used to estimate directly the distribution of WTP, and this in turn can be used to estimate the population mean. This claim is proved in McFadden (1994), which gives practical nonparametric estimators, and describes the restrictions necessary on referendum experimental design for these estimators to have good large-sample properties. In overview, the result is that with truthful referendum data there are estimators whose mean square error is inversely proportional to sample size, provided the experimental design "undersmooths" by taking a relatively large number of bid levels, with relatively small samples at each bid.² For example, when WTP is restricted *a priori* to a finite interval, one could distribute the bids evenly over this interval, with one respondent at each level. The common practice in CV referendum studies of taking a relatively small number of bid levels leads to estimators whose mean square errors decline more slowly with sample size.

² When the support of the WTP distribution is not finite, additional restrictions on tail behavior are needed to assure the existence of mean WTP and the stated rate of convergence of nonparametric estimators.

Conditional Versus Marginal Analysis

The estimation problem discussed above concentrates on the marginal distribution of WTP. This contrasts with most of the CV literature, which has emphasized estimation of conditional mean WTP given covariates that describe consumer characteristics. One usually collects covariates in a survey to check the representativeness of the sample and provide a basis for forming sample weights if needed. In many cases, the conditional relationship of WTP to covariates will be of direct interest as an aid to testing economic plausibility and comparing behavior across populations. However, the bottom line population unconditional mean WTP can be estimated at least as easily by the marginal methods in the previous paragraph as by a conditional approach that first estimates conditional mean WTP as a function of covariates, and then averages this with respect to an estimate of the density of the covariates. Conditional analysis is sometimes justified on the grounds that it is needed to correct nonrepresentativeness in the estimation sample. As I show below, when auxiliary information necessary to reweigh a sample is available, there may be statistical efficiency gains from a conditional approach. However, raking the estimation sample to make it representative and then applying marginal methods is consistent and often much simpler. I rephrase below a classical proposition on statistical sufficiency stating that when there is no information on the population distribution of the covariates other than that contained in the sample, then there is no statistical gain from conditional analysis. This proposition is followed by an example showing that conditional analysis can improve precision when there is auxiliary information on the covariates and sample sizes are large.

PROPOSITION. *Suppose there is a random sample with observations consisting of covariates, and a stated WTP if the experiment is open-ended, or a bid level drawn from a known density and a yes or no response if the experiment is single-referendum. Suppose WTP has its first three moments finite. Suppose there is no a priori or auxiliary information on the conditional distribution of the covariates, given the WTP data; i.e., the family of possible conditional distributions of the covariates is the same no matter what WTP data is observed. Then, for any estimator of WTP which may depend on the covariates, there is a second estimator which is a symmetric function of the*

marginal WTP data alone and is at least as precise in the sense of mean square error. Further, for open-ended observations, if there are no additional a priori restrictions on the marginal distribution of WTP, then the sample mean is a unique minimum variance unbiased estimator for mean WTP.

Proof. Let $p(\text{WTP} | \alpha)$ denote the marginal probability of the WTP data, with α denoting the population mean. Let $h(\mathbf{x} | \text{WTP})$ denote the conditional probability of the covariates. By assumption, h does not depend on α . Then, WTP is sufficient for α [Rao 2d2(c)]. Suppose $T(\text{WTP}, \mathbf{x})$ is any estimator of α . Form $T_1(\text{WTP}) = E[T(\text{WTP}, \mathbf{x}) | \text{WTP}]$; by sufficiency this statistic does not depend on α . Now, T_1 is uncorrelated with $T - T_1$, since $E T_1(T - T_1) = E_{\text{WTP}} T_1(E_{\mathbf{x} | \text{WTP}} T - T_1) = 0$. Hence, $E(T - \alpha)^2 = E(T - T_1 + T_1 - \alpha)^2 \geq E(T_1 - \alpha)^2$. This result is a version of the Rao-Blackwell theorem [Rao, 5a2(iii)]. The final result follows from an argument of Rao [5a2(ii)]: if $T(\text{WTP})$ is any unbiased estimator of α , then there is a second estimator formed by averaging over all permutations of the data that has no higher variance. Thus, there exists a minimum variance unbiased estimator that is symmetric in the observations. The sample mean is a symmetric unbiased estimator of α . If it were not minimum variance, there would exist a second unbiased estimator, which can be taken to be symmetric, with no higher variance. The difference in the two estimators would then be a symmetric unbiased estimator of zero. This would have to be true for any subfamily of the family of distributions with mean α , in particular the normal family. But the normal family admits no unbiased sufficient statistics for the population mean other than the sample mean, yielding a contradiction. This proves the proposition.

I next consider the example of a random sample of open-ended WTP data, and show that when there is auxiliary information on the covariates, there can be a gain in precision from a conditional approach. Suppose WTP W is jointly distributed with a vector of covariates \mathbf{x} that have mean μ and covariance matrix Σ . Consider the regression $W = \alpha + \mathbf{x}\beta + \varepsilon$, and let σ^2 denote the unconditional variance of the disturbance ε . Suppose that the distribution of (W, \mathbf{x}) is unknown, so that α , β , μ , Σ , and σ^2 are unknown. Suppose one has an independent auxiliary sample that yields an unbiased estimate $\hat{\mu}$ of μ . Let Ψ_N/N denote the covariance matrix of $\hat{\mu}$, scaled so that Ψ_N is comparable to the vari-

ance of one sample observation. Suppose the CV analyst adopts the variance-minimizing estimator $\hat{\mu} = (\Psi_N^{-1} + \Sigma^{-1})^{-1}(\Psi_N^{-1}\bar{\mu} + \Sigma^{-1}\bar{x})$ of μ . Then $\hat{\mu}$ has covariance matrix $(\Psi_N^{-1} + \Sigma^{-1})^{-1}/N$. The case $\Psi_N = 0$ corresponds to exact information on μ , and $\Psi_N = \infty$ corresponds to zero auxiliary information. The unconditional estimator for EW is the sample mean \bar{W} , which is unbiased and consistent, with variance $[\beta'\Sigma\beta + \sigma^2]/N$. The conditional estimator is $\hat{W} = a + \hat{\mu}\mathbf{b}$, the sample regression of W on one and \mathbf{x} , evaluated at $\hat{\mu}$. (It can also be written $\hat{W} = \bar{W} + (\hat{\mu} - \bar{x})\mathbf{b}$, implying that when there is no auxiliary information, $\hat{W} = \bar{W}$.) The covariance matrix of the conditional estimator is $[\beta'(\Sigma^{-1} + \Psi_N^{-1})^{-1}\beta + \sigma^2 + C_N]/N$, where $C_N = \mathbf{E}_{\{\mathbf{x}\}}\sigma^2(\hat{\mu} - \bar{x})(\Sigma(\mathbf{x} - \bar{x})'(\mathbf{x} - \bar{x})/N)^{-1}(\hat{\mu} - \bar{x})'$ is a term, arising from sampling variation in the estimator \mathbf{b} of β , that approaches zero as N increases. Then, when there is auxiliary information, there is a cross-over sample size above which the conditional approach has lower variance. When there is a single covariate that behaves such that the expectation defining $N \cdot C_N$ can be approximated by a probability limit, the cross-over sample size is approximately $(1 - R^2)/R^2$, where R^2 is the multiple correlation coefficient from the regression. These results do not require that \mathbf{x} be free of measurement error, or that the conditional mean of W be linear in \mathbf{x} , as failures simply contribute to the unconditional variance of the disturbance. Thus, the conditional approach is robust with respect to these modeling issues, and hence potentially broadly applicable. However, a major caveat is that the conditional approach is not robust with respect to inconsistencies between the sample and auxiliary distributions of covariates, such as differences in means due to differences in question wording, coding, or timing of the data collection.

The linear example above is perhaps a "best case" for conditional analysis. The nonlinear models required for referendum data are less robust with respect to functional assumptions, and typically require information on the complete joint distribution of the covariates, raising both practical computational problems and the statistical problem of the "curse of dimensionality" in nonparametric estimation. An exception that is amenable to conditional treatment is referendum data where all the covariates are categorical.

Outliers and Strategic Response

The estimation of population WTP is complicated if some sample responses are outliers, since a small number of consumers with very large stated WTP could substantially change an estimate of social value. While this is entirely appropriate if the extreme responses reflect genuinely high values for the resource, it leads to large sampling variances and nonrobustness with respect to measurement errors in WTP. The issue of statistical accuracy arises because the shape of a tail of a density is difficult to estimate. Nonparametric methods are inaccurate because of the difficulty of observing enough rare events to reliably estimate their frequency. Parametric models often pre-judge the issue by restricting the shape of the tail of the density, and are difficult to estimate precisely when they do allow flexible tail shapes. To illustrate, the hypothetical WTP densities $g_o(w) = w \cdot e^{-w}$ and $g_\varepsilon(w) = (1 - \varepsilon)g_o(w) + \varepsilon(1 + \varepsilon^2)(1 + w)^{-2-\varepsilon^2}$ have means 2 and $2(1 - \varepsilon) + 1/\varepsilon$, respectively. For $\varepsilon \leq 0.01$, the densities g_o and g_ε are virtually identical and would be impossible to discriminate with most samples, yet the mean of the second goes to infinity as ε approaches zero.

The problems of analyzing tail behavior are made far worse if survey respondents who strategically misrepresent their WTP tend to give extreme responses; e.g., some fraction of respondents treat the survey questions as an opportunity to express a protest for or against the preservation of the resource or the proposed vehicle of payment by giving a large magnitude response. Even a tiny fraction of consumers giving responses more extreme than their true WTP could lead to estimates of resource value that are in error by orders of magnitude; e.g., the distribution g_ε in the example above could be interpreted as a true WTP density $g_o(w)$ contaminated by strategic responses. This is a serious practical problem in CVM, as extreme responses in CV surveys are not uncommon, and responses to follow-up questions suggest that some subjects misinterpret the WTP elicitation or respond strategically.

The CV literature has attempted to address the problem of outliers and strategic misrepresentation by detecting and eliminating "protest" responses, by using statistical methods to estimate the mean that are less sensitive to extreme

observations, and by considering measures of value other than the population mean. Each approach introduces additional issues. Prescreening responses raises the question of what criteria should be used to identify suspicious responses, to classify suspicious responses as good or bad, and to handle observations judged to be bad. Classification errors in prescreening may cause subsequent statistical analysis to be inconsistent. The difficulty with the use of alternative statistical estimates of mean WTP, such as medians or trimmed means, is that they are themselves biased estimators when the distribution is skewed, as appears to be the case for many WTP distributions. Finally, social choice criteria other than Hicksian equivalent variation, such as median WTP, introduce some thorny issues centered on manipulation of the payment vehicle (McFadden 1994).

Parametric Analysis of CV Data

The estimators of mean WTP discussed above do not require any parameterization of the distribution of preferences. An alternative approach, widely used in CV analysis, is to specify a parametric model relating WTP to consumer income and other characteristics, and then to calculate the value of the commodity from the estimated model. Examples of parametric analysis of single-bid referendum data from CV studies are Cameron and James, and Bowker and Stoll, who specify probit models for referendum response: the probability of a yes response to a bid B when the consumer has a vector of covariates x is given by $\text{prob}(\text{yes} | x) = \Phi(x\beta - \alpha B)$; where Φ denotes the cumulative standard normal distribution, and β and α are parameters. Then, mean WTP is given by the formula $W = x\beta/\alpha$, which can be estimated using fitted parameters from the probit model.³

The advantages of parametric methods are that they make it relatively easy to impose preference axioms, combine experiments, and extrapolate the calculations of value to different populations than the sampled population. Their primary limitation is that if the parameterization is not flexible enough to describe behavior, then the misspecification will usually cause the mean WTP calculated from the estimated model

³ This formula for the mean holds when it is meaningful to have negative values of individual WTP, so that the distribution is symmetric. Alternately, if the WTP distribution is censored at zero, with a probability $\Phi(-x\beta)$ of a zero response, then mean WTP is $(x\beta/\alpha)\Phi(x\beta) + (1/\alpha)\phi(x\beta)$; and if the WTP distribution is truncated at zero, with zero probability of a nonpositive response, then mean WTP is $(x\beta/\alpha) + (1/\alpha)\phi(x\beta)/\Phi(x\beta)$.

to be an inconsistent estimate of true WTP. In particular, the estimated mean is highly sensitive to the assumption made on what parametric family contains the distribution of WTP, so that parametric estimates gain precision at the cost of decreased robustness.

I consider a flexible preference-based parametric model for WTP W that includes some of the commonly used models in the CV literature

$$(1) \quad W = W(y, z; \alpha) \\ \equiv y - [y^{1-\alpha} - (1-\alpha)z]^{1/(1-\alpha)}$$

where y is real discretionary income and z is an unobserved (latent) variable expressing tastes for the environmental resource. In this expression, α is a nonnegative parameter that can be interpreted as the "income elasticity" of WTP.⁴ At $\alpha = 1$ the formula reduces to $W = y(1 - e^{-z})$. The taste variable z is restricted to the range $-\infty < z \leq z_T = y^{1-\alpha}/(1-\alpha)$ for $\alpha < 1$, and to the range $z_T < z < +\infty$ for $\alpha > 1$. Formula (1) is derived from an indirect utility function

$$V(W, z) = z + \left[\frac{(y - W)^{1-\alpha} - 1}{1 - \alpha} \right]$$

that is a "Box-Cox" transformation of net income plus the contribution z from the existence of the resource. The equivalent variation W equates utility with the resource to utility without, $V(y - W, z) = V(y, 0)$.⁵ Additional discus-

⁴ A Taylor's expansion of (1) around $z = 0$ shows that for positive w that is less than 10% of income, the elasticity of w with respect to y is virtually identical to α , and the slope of the WTP function with respect to z is virtually one.

⁵ This indirect utility function can be derived from a CES direct utility function of private goods and public goods, with z determined by the levels of public goods with and without the resource. The indirect utility function can be generalized to make discretionary income a parametric function of total income; and, for commodities with a use component, to append an additive term that depends on the user price; see McFadden and Leonard. If $\alpha < 1$, there is for $z > z_T$ a corner solution where the consumer has higher utility with than without the resource, even when all income is committed to its preservation. If $\alpha > 1$, there is a corner solution for $z < z_T$ where the consumer has higher utility without than with the resource, even when all income is committed to its removal. If $\alpha = 1$, the equivalent variation is defined for all z . There is a philosophical issue of whether an individual can rationally offer a payment to preserve or remove a resource that would violate conditions for her own subsistence, thus existence, and if so how her preferences should be treated in calculating social welfare. In further analysis, I circumvent this issue by assuming that zero income is the threshold for subsistence, and that preferences are lexicographic at subsistence so that WTP is always defined, if necessary, by a corner solution. Most CV practitioners would treat a survey response with stated WTP above (a substantial fraction of) income as a strategic response that misrepresents true preferences, so the corner problem would not appear in the accepted data.

sion of the qualitative properties of W that follow from utility theory can be found in McConnell, and McFadden and Leonard. The taste variable z is assumed to have a cumulative distribution function $F(z)$, which may depend on covariates \mathbf{x} such as age or education if they are present, and will depend on parameters when F is restricted to a parametric family.⁶

The cumulative distribution function of an open-ended response w for $w < y$ is

$$(2) \quad \text{prob}(W \leq w) \equiv F\left[\frac{y^{1-\alpha} - (y-w)^{1-\alpha}}{1-\alpha}\right];$$

the probability is one for $w \geq y$. In some applications, negative w may be excluded by assumption or by survey design; then, a positive probability of $w = 0$ can be accommodated by specifying a family for which $F(0) > 0$ is possible, with a level that can be determined by survey responses.⁷

The probability for a single-bid referendum response is

$$(3) \quad \text{prob}(no \mid B) = F\left[\frac{y^{1-\alpha} - (y-B)^{1-\alpha}}{1-\alpha}\right]$$

where B is the bid. The probability for a double-bid referendum response, when the bids are B' and B'' with $B' < B''$, is

$$(4) \quad \text{prob}(both \ no \mid B', B'') = F\left[\frac{y^{1-\alpha} - (y-B')^{1-\alpha}}{1-\alpha}\right]$$

⁶ The function F is a description of preferences that are determined prior to specification of the consumer's budget. Hence, \mathbf{x} cannot include income or prices. In practice, misspecification of the utility function, or correlation of income or prices with unobserved covariates that influence tastes, may lead to an apparent conditioning of F on income or prices. However, it would be logically inconsistent to include these variables in \mathbf{x} and carry through welfare calculations in which income or price changes alter the distribution F .

⁷ Censoring at z_p is a consequence of the assumptions on utility maximization. Then, in using the distribution of z to generate a distribution of W in the case $\alpha < 1$, right-censoring with a point mass of probability at z_p should be used. Censoring at zero may be due to behavior, or to a survey format that does not permit negative values of W . In the former case, the point mass of probability at zero should be used. In the latter case, the distribution should not be left-censored when the distribution of W is generated or mean WTP is calculated. A CV experiment that fails to discriminate zero and negative values of WTP cannot determine what combination of these cases is applicable in the calculation of social value.

$$(5) \quad \text{prob}(bracket \mid B', B'') = F\left[\frac{y^{1-\alpha} - (y-B'')^{1-\alpha}}{1-\alpha}\right] - F\left[\frac{y^{1-\alpha} - (y-B')^{1-\alpha}}{1-\alpha}\right]$$

$$(6) \quad \text{prob}(both \ yes \mid B', B'') = 1 - F\left[\frac{y^{1-\alpha} - (y-B'')^{1-\alpha}}{1-\alpha}\right]$$

An implication of the rational preference model is that the probability of a bracket, conditioned on bids B' and B'' , is independent of whether the higher or lower bid is presented first.

When the taste distribution F is restricted to a parametric family, the parameters of the distribution and the parameter α of the indirect utility function can be estimated by maximum likelihood from either open-ended or referendum data.

The family (1) can be specialized to two models that have traditionally been used in the CV literature; if $\alpha = 0$, there is no income effect on the WTP and one has $\text{prob}(W \leq w) = F(w)$ for $w < y$. Then, WTP can be described by a doubly censored "regression" model

$$w = \begin{cases} \beta + \xi & \text{if } -\beta \leq \xi \leq y - \beta \\ y & \text{if } \xi > y - \beta \\ 0 & \text{if } \xi < -\beta \end{cases}$$

where ξ is a random variable with expectation zero and β is the mean of the F distribution; the censoring at zero is needed only if survey observations are so censored. Mean WTP in the population is given by the mean of the F distribution adjusted for censoring at $w = y$ and, if appropriate, at $w = 0$.

If $\alpha = 1$, then $\text{prob}(W \leq w) = F[-\log(1 - w/y)]$, and one has a censored regression model

$$-\log(1 - w/y) = \begin{cases} \beta + \xi & \text{if } -\beta < \xi \\ 0 & \text{if } \xi \leq -\beta \end{cases}$$

where β is the mean of the F distribution; the censoring is needed in estimation only if the survey responses are censored at zero.

Table 1. Parametric Taste Distributions

Family	Distribution function	Parameters	Restrictions	Mean before censoring
Normal	$\Phi\left(\frac{z - \beta}{\sigma}\right)$ for $z \geq 0$	β, σ	$\sigma > 0$	β
Log normal	$\begin{cases} 1 - \pi & \text{for } z = 0 \\ 1 - \pi + \pi\Phi\left(\frac{\log(z) - \beta}{\sigma}\right) & \text{for } z > 0 \end{cases}$	π, β, σ	$0 \leq \pi < 1$ $\sigma > 0$	$\pi \cdot \exp(\beta + \sigma^2/2)$
Gamma	$\begin{cases} 1 - \pi & \text{for } z = 0 \\ 1 - \pi + \pi \int_0^z \frac{u^{\beta-1} e^{-u}}{\Gamma(\beta)} du & \text{for } z > 0 \end{cases}$	π, β, λ	$0 \leq \pi < 1$ $\lambda > 0$ $\beta > 0$	$\pi\beta/\lambda$
Weibull	$\begin{cases} 1 - \pi & \text{for } z = 0 \\ 1 - \pi \cdot \exp\left[-(z/\beta)^\lambda\right] & \text{for } z > 0 \end{cases}$	π, β, λ	$0 \leq \pi < 1$ $\lambda > 0$ $\beta > 0$	$\pi\beta\Gamma(1+1/\lambda)$

Hanemann (1984) has employed this form in CV regressions.

For parametric analysis, I consider four families F for the distribution of unobserved tastes: normal, with censoring at zero; log-normal; gamma; and Weibull. Each of the last three distributions is mixed with a point mass at zero. Table 1 gives some features of these families. These distributions are defined without covariates. When I introduce consumer characteristics \mathbf{x} , I make the parameter β in these models a function of these covariates. In the normal or log normal, I replace β by $\beta_0 + \mathbf{x}\beta_1$, and in the gamma and Weibull, I replace β by $\beta_0 \exp(\mathbf{x}\beta_1)$, where β_1 is a vector of parameters.

Experiment

To test the properties of CVM for an existence value problem, I have analyzed experiments in which respondents are asked to value a resource by either an open-ended (OE) response or a referendum response with two bids. The first response to the double referendum (DR) also provides information on one-bid referendum response (SR). Independent samples are used for each elicitation format and experimental treatment, for several resource valuation

problems that are related by the scale of the resource and by the information provided to the respondents. The resource issue selected for the experiment is described in the following introductory paragraph that was used in most of the experimental treatments: "In the states of Colorado, Idaho, Montana, and Wyoming there are fifty-seven federally protected 'wilderness areas' with a total of 13 million acres. They are managed by the United States Forest Service and are similar to national parks, except that roads, commercial development, mechanical equipment, and other improvements are prohibited. Access is limited to such personal uses as hiking, camping, fishing, and hunting. These 'wilderness areas' were established under the 1964 Wilderness Act, which was designed to preserve the natural environment. While these areas vary in size, from less than 50,000 acres to more than 2 million acres, they are quite similar in terms of the habitats they provide and the remoteness of their locations. The Selway Bitterroot Wilderness in northern Idaho is one of these fifty-seven areas. It covers 1.3 million acres, including parts of four national forests, between Idaho's Salmon River and the Montana border. It is one of the largest 'wilderness areas' in the continental United States. In light of the federal budget deficit, it has been proposed

Table 2. Experimental Treatments

Open-ended response	Double referendum response
OE1. Household WTP for Selway	DR1. Household WTP for Selway
OE2. Household WTP for all 57 areas	DR2. Household WTP for all 57 areas
OE3. Household WTP for Selway, no information on other areas	DR3. Household WTP for Selway and two other areas
OE4. Individual WTP for Selway, no information on other areas	DR4. Household willingness to vote for Selway

that the federal government lease the Selway Bitterroot Wilderness for commercial development as a way of raising revenue. One proposal for commercial development involves allowing timber companies to harvest the mature timber at a rate of 1% per year, indefinitely. This would necessitate building roads and bringing in mechanical equipment. An alternative way the government might increase revenues, without disturbing the Selway Bitterroot Wilderness, would be through a federal income tax surcharge designated for wilderness preservation. We are interested in the value that your household places on making certain the Selway Bitterroot Wilderness is preserved.”

This resource issue was chosen because at the time of the study in 1990 there was active discussion in Congress and in the media in the western U.S. regarding logging on government lands, making the topic of logging familiar to many respondents. Further, the remoteness and limited accessibility of the various wildernesses, and the availability of substitutes, implied that the value that respondents placed on the resource would be predominantly existence rather than active use value. The samples were drawn randomly from households in four states: Colorado, Idaho, Montana, and Wyoming. This population was chosen so that a majority of respondents were likely to have some knowledge of the Selway-Bitterroot Wilderness and other wilderness areas, but unlikely to have utilized the Selway area. In the combined experiments, about 71% of the respondents had heard of the Selway area before the survey, and 24% had visited the area or its fringes. The subjects were contacted by random digit dialing from a population defined as the heads of household at least eighteen years of age in the four states. Selected telephone numbers were retried at least ten times at different times of day and days of the week if there was no answer. Overall, the

response rate for the surveys averaged 62%.⁸ Following standard practice in CV analysis, the responses were screened to remove “protest zeros”; individuals giving a zero WTP were asked if this was because they did not value the commodity, or because they objected to the payment vehicle or some other aspect of the question. Respondents were excluded from the sample if they stated that their zero response was due to an objection. There was no screening of positive responses. Final sample size was 1,229.⁹

Table 2 describes the experiments. The effect of elicitation format can be tested by comparing OE1 with DR1, or OE2 with DR2. The effect of scope of the preservation effort can be tested by comparing OE1 with OE2, or DR1 with DR2. In order to test some additional aspects of survey methodology, some of the experiments differed between the open-ended and closed-ended designs. Experiments OE1 and OE3 differ in that the existence and status of wildernesses other than Selway were not mentioned in OE3. They can be compared to test for the framing effect of information provided to the subject. Experiments OE3 and OE4 differ only in that they ask for household and individual WTP, respectively. Their comparison provides informa-

⁸ One referee judges the resource issue to be implausible and imprecisely framed. Such flaws would tend to exacerbate psychometric biases. However, based on my experience with preference measurement in marketing and psychology, I would expect qualitatively similar response patterns in even the most topical CV studies. This referee also objects to the use of telephone interviews (CATI). I know no evidence that psychometric biases are more severe in telephone surveys than in personal surveys. While it is a severe limitation to not use visual materials, card sorts, and other procedures available in home interviews, there are great advantages to CATI in cost, monitoring of interviewers, sample frame, and response rates. CV practitioners would do well to look very carefully at performance/cost ratios for alternative survey methods before ruling out CATI.

⁹ The data used in this study are available by e-mail from the author; send requests to mcfadden@econ.berkeley.edu.

tion on the imputations respondents make for other household members. Experiment DR3 considers an intermediate scope of preservation. Experiment DR4 differs from DR1 only in the wording "agree to pay" is replaced by "vote for."

The open-ended questions were phrased as follows: "If you could be sure the Selway Bitterroot Wilderness would not be opened for the proposed TIMBER HARVEST and would be preserved as a 'wilderness area' indefinitely, what is the MOST that your HOUSEHOLD would pay EACH YEAR through a federal income tax surcharge, designated for preservation of the SELWAY BITTERROOT WILDERNESS?"

The referendum questions were asked for a randomly assigned first bid chosen from the following levels: \$2, \$10, \$20, \$40, \$60, \$100, \$200, \$1,000, \$2,000. The phrasing of the question follows: "If you could be sure the Selway Bitterroot Wilderness would not be opened for the proposed TIMBER HARVEST and would be preserved as a 'wilderness area' indefinitely, would your HOUSEHOLD agree to pay a federal income tax surcharge of \$B per year, designated for preservation of the SELWAY BITTERROOT WILDERNESS?"

A yes response was followed up by the same question with twice the first bid, and a no response was followed up by the same question with half the first bid. Subjects were not told in advance that there would be a follow up bid, so first-response data can be treated as a single-referendum experiment; these are denoted SR1 to SR4.

Results

The strategy for analysis that I follow is to concentrate on the parametric models previously described, and then to check the specification using nonparametric methods. I first consider the specification of the taste distribution. I next test for common income effects, and test two special cases that are common in the CV literature. After this, I examine the "embedding effect" of the scope of the resource. I follow by examining the context effects of question phrasing and information provided. Finally, I examine the comparability of open-ended and referendum responses.

Distribution of Tastes

The parametric model of WTP previously described was estimated by maximum likelihood for each of the experiments OE1–OE4 and

DR1–DR4, using each of the taste distributions described in table 1. The likelihood function was formed using equation (2) for the open-ended experiments, equation (3) for the single-referendum experiments (that ignore the second response), and equations (4)–(6) for the double-referendum experiments. In each experimental treatment, the mixed log-normal taste distribution gives the highest likelihood, and is also the preferred distribution using the Akaike information criterion. In the referendum experiments, the mixed Weibull distribution is a close second. In all cases, the censored normal distribution is substantially inferior to the remainder.

Parametric Restrictions and Covariates

An implication of the parametric specification (1) is that there should be a common income effect across experimental treatments. To test this, I estimate the model with the mixed log normal distribution of tastes, imposing a common income parameter α , and allowing the remaining parameters to vary by experiment. These estimates for each elicitation format are given in table 3. The first three panels of the table give the results for the open-ended, single-referendum, and double-referendum elicitation formats, respectively. The final panel gives results when a common value of α is imposed on the pooled open-ended and double-referendum experiments. The third and fourth columns in this table give the maximized log likelihood and estimate of α for each experiment fitted separately, and also these statistics when a common α is imposed within each elicitation format. The following columns of the table give estimates of the remaining parameters for each experiment obtained when a common α is imposed for each treatment. Using likelihood ratio tests, I accept (at the 10% level) the hypothesis that there is a common α within each elicitation format.¹⁰ The test of a common α between the open-ended and either referendum format can be done using a t-statistic or a likelihood ratio statistic; either procedure accepts (at the 10% level) the hypothesis of a common α across elicitation formats. The t-statistic for the independent samples OE versus SR is 0.173, and for the independent

¹⁰ The LR test statistics for a common α across the experiments within OE, SR, and DR are 3.450, 2.152, and 1.820, respectively. Each is chi-square with three degrees of freedom under the null, with a 10% critical level of 6.125.

Table 3. Parametric Models of WTP

Experiment	Sample size	Log likelihood	Alpha	Beta	Sigma	Pi	Mean WTP	Median WTP
OE1	256	-1014.410	0.253 0.137	0.077 0.832	1.499 0.082	0.688 0.029	44.65 8.42	8.28 2.62
OE2	275	-1246.480	0.408 0.162	0.682 0.835	1.344 0.062	0.735 0.027	70.09 11.37	19.98 5.42
OE3	279	-1056.220	0.096 0.143	-0.149 0.821	1.543 0.088	0.681 0.028	37.74 6.80	6.22 2.10
OE4	72	-262.195	0.716 0.343	-0.220 0.849	1.414 0.164	0.667 0.056	28.46 9.06	5.85 2.63
OE Pooled	882	-3581.030	0.290 0.081	na	na	na	na	na
SR1	365	-225.880	-0.090 0.354	0.184 3.120	4.453 1.949	1.051 0.455	3256.23 1460.77	41.67 27.24
SR2	365	-222.955	0.424 0.342	2.215 1.754	3.000 0.800	0.847 0.104	2880.12 1234.59	121.34 74.42
SR3	376	-211.562	0.560 0.291	1.303 1.734	2.477 0.459	0.891 0.096	1063.17 452.49	66.06 36.22
SR4	368	-222.228	0.221 0.342	2.328 1.712	1.717 0.388	0.618 0.058	664.15 305.81	59.42 37.43
SR Pooled	1474	-883.701	0.321 0.163	na	na	na	na	na
DR1	365	-440.084	0.203 0.205	0.040 1.025	2.353 0.219	0.909 0.051	536.47 141.10	36.08 14.38
DR2	365	-471.106	0.333 0.187	0.795 1.006	2.118 0.169	0.901 0.034	717.47 178.83	77.03 29.18
DR3	376	-397.443	0.587 0.203	0.323 1.034	1.925 0.172	0.883 0.043	346.44 85.65	46.75 16.89
DR4	368	-470.017	0.362 0.201	0.182 1.020	2.087 0.194	0.837 0.042	372.28 109.91	33.32 12.73
DR Pooled	1474	-1779.560	0.377 0.098	na	na	na	na	na
OE1	na	na	na	-0.273 0.548	1.500 0.082	0.688 0.029	44.92 8.22	8.25 2.39
OE2	na	na	na	0.329 0.649	1.343 0.061	0.735 0.027	70.09 10.89	19.87 4.80
OE3	na	na	na	-0.496 0.639	1.546 0.089	0.681 0.028	38.16 6.72	6.21 1.91
OE4	na	na	na	-0.573 0.670	1.410 0.163	0.667 0.056	28.95 8.83	5.83 2.51
DR1	na	na	na	0.579 0.684	2.351 0.219	0.909 0.052	529.70 137.74	36.00 10.75
DR2	na	na	na	1.335 0.654	2.115 0.168	0.901 0.034	707.65 176.41	76.81 20.86
DR3	na	na	na	0.869 0.675	1.927 0.174	0.884 0.044	345.81 84.89	46.90 12.34
DR4	na	na	na	0.727 0.666	2.086 0.194	0.836 0.042	369.11 108.06	33.38 9.45
OE/DR Pooled	2356	-5360.83	0.324 0.063	na	na	na	na	na

Mixed log-normal model, maximum likelihood estimates, standard errors below estimates.

samples OE versus DR is 0.688. Alternately, the likelihood ratio statistic for a common α across all OE and DR experiments is 5.75; the 10% critical level for this statistic, with seven degrees of freedom, is 12.02. The SR and DR samples are common, with more information in the DR data under the hypothesis, so that one can form a Hausman statistic, given by the

square of the difference of the parameter estimates divided by the difference of their estimated variances. This statistic is 0.187; the 10% critical level, with one degree of freedom, is 2.71. A common maximum likelihood estimate of α , formed by pooling the open-ended and double-referendum data, is 0.324 (with a standard error of 0.063). These tests are not

powerful enough to rule out every economically significant variation in α , but with this caveat the results strongly support the parameter specification (1).

The last two columns of table 3 give estimates of the population mean and median WTP, and estimates of their standard errors. From (1) and the mixed log-normal distribution for the taste variable z , mean WTP is estimated by

$$\bar{E}W = \sum_y \hat{p}(y)H(y, \hat{\theta})$$

where $\hat{p}(y)$ is the pooled sample estimate of the (multinomial) distribution of income over response brackets; $\theta = (\alpha, \beta, \sigma, \pi)$ is the vector of parameters estimated by maximum likelihood with α restricted to a common value within the elicitation format; and $H(y, \theta)$ is the function (for $\alpha < 1$)

$$H(y, \theta) = \pi \left\{ y - \int_{-\infty}^L [y^{1-\alpha} - (1-\alpha)e^u] \phi\left(\frac{u-\beta}{\sigma}\right) \frac{du}{\sigma} \right\}$$

with $L = \log y^{1-\alpha}/(1-\alpha)$. The variance of the estimate $\bar{E}W$ is obtained by the delta method, and is approximately

$$\left[\sum_y \hat{p}(y)H_{\theta}(y, \hat{\theta}) \right] V(\hat{\theta}) \left[\sum_y \hat{p}(y)H_{\theta}(y, \hat{\theta}) \right] + H'V(\hat{p})H$$

where H_{θ} is the vector of derivatives of H with respect to the parameters, H is the vector with components $H(y, \hat{\theta})$ for each y , $V(\hat{\theta})$ is the covariance matrix of the maximum likelihood estimates, and $V(\hat{p})$ is the covariance matrix of the multinomial estimates of the income distribution. The estimate of the population median is the solution \hat{M} to the equation

$$0.5 = \sum_y \hat{p}(y)J(y, \hat{\theta}, \hat{M})$$

where

$$J(y, \theta, M) \equiv \Phi \left(\frac{\log \left[\frac{y^{1-\alpha} - (y-M)^{1-\alpha}}{1-\alpha} \right] - \beta}{\sigma} \right)$$

The estimated variance of the estimate of the median is also approximated using the delta method, and is

$$\left[\sum_y \hat{p}(y)J_M(y, \hat{\theta}, \hat{M}) \right]^2 \left\{ \left[\sum_y \hat{p}(y)J_{\theta}(y, \hat{\theta}, \hat{M}) \right] V(\hat{\theta}) \left[\sum_y \hat{p}(y)J_{\theta}(y, \hat{\theta}, \hat{M}) \right] + J'V(\hat{p})J \right\}$$

where J_M is the derivative of J with respect to M , J is the vector with components $J(y, \hat{\theta}, M)$ for each y , and the remaining terms have the same definition as in the case of the estimated mean. The formulas giving the population mean and median estimates are calculated by numerical integration.

It is common practice in CV analysis to use the specification (1) and impose the special case $\alpha = 0$ or $\alpha = 1$. The first case implies that $W = z$ has the same distribution as z , while the second case implies that the transform $-\log(1 - W/y)$ has the same distribution as z . I use t-statistics to test the hypotheses $\alpha = 0$ and $\alpha = 1$ in each elicitation format, maintaining the hypothesis of a common value of α for the experiments within a format. These hypotheses are rejected at the 5% level. I conclude that these specializations are not consistent with behavior, and note that some inconsistencies in CV results in the literature could arise from this misspecification of the income effect. An economic interpretation of the results on the relationship of income to WTP in these experiments is that preservation of the Selway-Bitterroot wilderness is a "necessary" good, with a low income elasticity. However, it seems economically plausible that preservation would be a "luxury" good that for poor households is displaced by needs for food and shelter. Reporting errors and grouping will reduce the magnitude of estimated income effects, but the difference here is large enough to raise the question of whether responses are nonstrategic and reflect stable rational preferences.

For further testing of the parametric specification, I consider the null hypothesis that the location of the log-normal taste distribution is a constant β_0 , against the alternative that it is a linear function $\beta_0 + \beta_1x$ of a vector of covariates x . First, I test the parametric dependence of WTP on income implied by model (1). I consider the alternative in which β_1x is a quadratic function of income, with parameters that can differ for each of the eight experiments.

Table 4. Effect of Covariates on WTP

Specification issue	Experiment							
	OE1	OE2	OE3	OE4	DR1	DR2	DR3	DR4
Age and sex								
Age	-0.018	-0.017	-0.025	-0.014	-0.067	-0.052	-0.036	-0.053
	0.010	0.007	0.008	0.016	0.012	0.010	0.010	0.011
Sex	0.270	0.022	-0.018	-0.023	0.986	0.563	0.517	0.098
	0.223	0.190	0.222	0.415	0.323	0.297	0.299	0.313
Demographics								
Age	-0.108	-0.017	-0.026	-0.011	-0.069	-0.051	-0.033	-0.049
	0.010	0.075	0.086	0.016	0.013	0.011	0.011	0.012
Sex	0.267	-0.017	-0.008	0.029	0.972	0.648	0.492	0.123
	0.222	0.198	0.222	0.478	0.322	0.303	0.299	0.314
Education	0.037	-0.069	-0.045	-0.088	0.037	0.145	0.062	0.034
	0.055	0.038	0.040	0.110	0.056	0.060	0.058	0.065
log(household size)	-0.269	-0.050	-0.128	0.221	-0.386	0.019	0.206	0.302
	0.226	0.211	0.228	0.414	0.338	0.258	0.305	0.338
Use value effects								
Outdoor activities	0.043	0.025	0.212	0.081	0.595	0.082	0.127	0.494
	0.133	0.120	0.156	0.281	0.214	0.187	0.173	0.200
Location	0.431	0.450	0.669	0.083	-0.165	-0.638	-0.176	-0.150
	0.262	0.222	0.270	0.437	0.375	0.320	0.329	0.339
Visit area	0.645	0.347	0.022	1.123	0.411	0.066	-0.533	0.316
	0.315	0.227	0.280	0.864	0.442	0.383	0.338	0.477
Summary Statistics:								
	Likelihood under null	Likelihood under alt.	LR statistic	DF	Critical level	Signif. level	Sample size	Alpha & SE
Age and sex	-5346.17	-5284.92	122.50	16	32.00	1%	2349	0.344 0.063
Demographic effects	-5284.92	-5275.34	19.16	16	23.54	10%	2349	0.332 0.070
Use value effects	-5277.27	-5248.41	57.72	24	42.98	1%	2321	0.315 0.064

(Mixed log-normal model, maximum likelihood estimates, standard error given below each estimate.)

This alternative does not correspond in general to utility maximization, but nevertheless should have power against monotone dependence of WTP on income that is not captured by (1). The hypothesis that (1) adequately describes income effects is accepted at the 10% level using a likelihood ratio test (LR = 14.54, DF = 16, critical level = 23.54).

In giving the parameter α an economic interpretation, note that the unconditional analysis will attribute to income the effects of any covariates, such as age or education, that are correlated with income. This is not an issue in forecasting mean WTP, as the conditional distribution of WTP given income, weighted by the empirical distribution of income, will still yield a consistent estimate of the population mean. However, these correlates can confound the behavioral interpretation of income effects. To test for the presence of confounding effects,

I carry out a specification test with age and sex as covariates entering β_1x . Table 4 gives the estimated coefficients of age (in years) and sex (dummy for female head of household) for each experiment. One sees that age has a consistent negative effect on WTP; this may be a true age effect, or may be a cohort effect. Sex of respondent does not have a significant effect on tastes in the OE experiments, but does appear to increase WTP in the DR experiments. This result suggests that females may be more sensitive than males to the contextual information provided by referendum bid levels.

There is a modest positive correlation of age and income in the sample, so that some of the observed income effect might be due to an age effect. However, the estimate of α when age and sex are included as covariates is virtually unchanged. Table 4 also gives a test for the effect of education (in years) and log(household

size) as covariates, given the presence of age and sex, and finds that these variables do not influence tastes.

Experiments OE3 and OE4 differ only in that OE3 asks for household WTP, and OE4 asks for individual WTP. If the distribution of WTP for the household heads that are survey respondents is the same as the distribution of WTP in the population as a whole, and when asked for household WTP the respondents treat all household members equally, then the household response should on average equal the individual response multiplied by household size.¹¹ In the range of the observations and fitted parameters, $\log(\text{WTP})$ is virtually linear in $\log(\text{household size})$. Therefore, if individual WTP is independent of household size, and reported household WTP is proportional to household size, then the coefficient of this variable should be one for OE3 and zero for OE4. Even if WTP is not independent of household size, the coefficient in OE3 should equal the coefficient in OE4 plus one if respondents distinguish personal WTP from the WTP they impute to the household. Using Wald tests, I accept at the 10% level the hypothesis that the coefficient is zero in both OE3 and OE4 (statistic = 1.04, DF = 2, critical level = 4.61), and reject at the 1% level the hypothesis that the coefficient in OE3 equals one plus the coefficient in OE4 (statistic = 8.15, DF = 1, critical level = 6.63). I conclude that consumers do not attribute to other household members a WTP as high as their own. The estimates of mean WTP give the result that WTP for households is 31.8% higher than WTP for individuals. Mean household size is 2.6; this implies that the mean WTP imputed to each other household member by the head is \$5.68, in contrast to individual mean WTP of \$28.95 for the head. While it is plausible that consumers may treat children as less than full adult equivalents, this difference gives the implausible implication that WTP of spouse is less than 32% of WTP of head. This suggests that many consumers fail to distinguish between questions asking for household or individual response, that they do not form their responses by aggregating linearly over household members, or that there is bias in their imputation of the WTP of others. One possible explanation is that questions on individual WTP elicit responses that already incorporate an imputation of value to other household members or associates, per-

haps because the survey preamble focused on the household unit.

While it is arguable that consumer value for the Selway-Bitterroot Wilderness is primarily for existence, it is possible that there is a user component for some consumers. To test this, I consider a specification in which β_1x contains the following variables: "outdoor activities," a count of the number of activities among camping, hiking, fishing, and hunting the household currently or previously engaged in; "location," a dummy variable for state residence in Idaho or Montana, the areas near Selway; and "visit," a dummy variable which is one if the household has previously visited the area being evaluated. A likelihood ratio test finds that these variables are jointly significant. Table 4 gives the estimated coefficients. Number of outdoor activities appears to have no effect on OE responses, but appears to have a positive effect on WTP for some of the DR experiments. Location near Selway has a significant negative effect on WTP in the open-ended experiments OE1 and OE3 that value this area alone, and an insignificant positive effect in the referendum experiments DR1 and DR4 that value Selway. A possible explanation for the negative effect in the OE experiments is that residents of Idaho and Montana on average place lower value on specific wilderness areas than residents of Wyoming and Colorado, irrespective of potential user benefits, perhaps due to greater availability of substitutes, or perhaps due to greater participation in economic benefits from increased logging activity. The failure of a negative proximity effect to appear in the DR experiments suggests that this format may lead respondents to concentrate more on the benefits of preservation, and less on the opportunity costs. The coefficient of "visit" is positive for all except one experiment, but statistically insignificant in all but one case. I conclude that the experimental valuations of wilderness preservation include some use value, but that existence value is the main component of stated total value.

Embedding

A phenomenon that has been found in a number of CV studies is an embedding effect, in which the value placed on a resource is virtually independent of the scale of the resource (Kahneman and Knetsch). While some decline in marginal value with scale would be expected in classical consumer theory due to substitution, with an

¹¹ I see no reason to assume that existence values share the economies of scale and consumption-in-common that motivate household equivalent scales.

additional contribution from income effects if values are substantial relative to income, the observed embedding effects are far too large in magnitude to be explained plausibly using classical preference theory. In the Selway experiments, DR1, DR2, and DR3 differ only in the scope of the resource being valued, with DR1 asking for the value of the Selway Bitterroot Wilderness, DR3 asking for the value of this area plus two other wilderness areas (Washakie and Bob Marshall), and DR2 asking for the value of all fifty-seven wilderness areas. Similarly, OE1 asks the value of Selway Bitterroot and OE2 asks the value of all fifty-seven areas. Then, these experiments can be used to test for the presence of the embedding effect.

First for the OE experiments, and then for the DR experiments, I test the hypothesis that the population mean value is independent of resource scale, and the hypothesis that mean value is proportional to resource scale.¹² A t^2 -statistic, which is asymptotically chi-square with one degree of freedom under the null, is used for the OE tests, while the DR test for scale-independent value uses the Wald statistic

$$\begin{bmatrix} \hat{\mu}_3 - \hat{\mu}_1 \\ \hat{\mu}_{57} - \hat{\mu}_1 \end{bmatrix}' \begin{bmatrix} V_3 + V_1 & V_1 \\ V_1 & V_{57} + V_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mu}_3 - \hat{\mu}_1 \\ \hat{\mu}_{57} - \hat{\mu}_1 \end{bmatrix}$$

where $\hat{\mu}_k$ is the estimated mean WTP for k areas, and V_k is the estimated variance of this estimator. This statistic is asymptotically chi-square with two degrees of freedom when the null hypothesis is true. In the tests for values proportional to scale, $\hat{\mu}_k$ is replaced by $\hat{\mu}_k/k$, and V_k by V_k/k^2 . The null hypothesis that value is independent of scale between OE1 and OE2 is accepted at the 5% level (t^2 statistic = 3.233, critical level = 3.84), and between DR1, DR2, and DR3 is accepted at the 10% level (Wald statistic = 3.957, critical level = 4.61). The null hypothesis that WTP is proportional to scale between OE1 and OE2 is rejected at the 1% level (t^2 statistic = 26.58, critical level = 6.63); and between DR1, DR2, and DR3 is also rejected at the 1% level (Wald statistic = 27.057, critical level = 9.21). Thus, these experiments demonstrate the embedding effect. The parameter estimates from table 3 indicated that stated WTP from DR3 for three areas including Selway is 35.4% lower than the value

of Selway alone from DR1, and WTP for fifty-seven areas is only 33.8% higher than WTP for Selway alone. WTP for fifty-seven areas from OE2 was 57% higher than WTP for Selway alone in OE1; strict proportionality would require that it be 5,700% higher. These results indicate that either there are extraordinarily strong diminishing returns to preserving additional wilderness areas, or that there is a context effect that makes responses inconsistent with classical economic preferences. Analyzing a series of closely related experiments with alternatives designed to control for classical diminishing marginal utility, Diamond, Hausman, Leonard, and Denning conclude that the embedding effect is inconsistent with classical preferences over resource allocations.

Information and Phrasing

Experiments DR1 and DR4 differ only in the phrasing of the bid, with DR1 using the phrase "agree to pay" and DR4 using the phrase "vote for." If consumers have stable preferences, and responses are revealing preferences rather than strategic misrepresentation motivated by the perceived incidence of the payment vehicle, then the responses should be insensitive to this change in context. A t -statistic for differences in estimated means using first responses is 1.74, and using both first and second responses is 1.06. Then, the hypothesis that response is insensitive to this variation in wording is accepted at the 5% level. Although the difference is not statistically significant, the "agree to pay" wording yields a 44.1% higher mean WTP than the "vote for" wording. Thus, the results are ambiguous, but do provide at least weak support for the proposition that this minor phrasing change does not have a substantial effect on stated WTP.

Experiments OE1 and OE3 both ask for WTP for Selway. They differ only in the information provided about other wilderness areas, with OE3 not mentioning that other wilderness areas are threatened by logging. The null hypothesis is that the valuation of Selway is insensitive to this information. A t -statistic for this hypothesis is 0.64, so the null is accepted at the 10% level.

Sensitivity to Elicitation Format

I test the hypothesis that the distribution of WTP obtained from open-ended and referendum responses is the same. If behavioral re-

¹² The wildernesses differ in size and features, and Selway-Bitterroot is one of the largest, so that exact proportionality is unlikely even if there is no embedding.

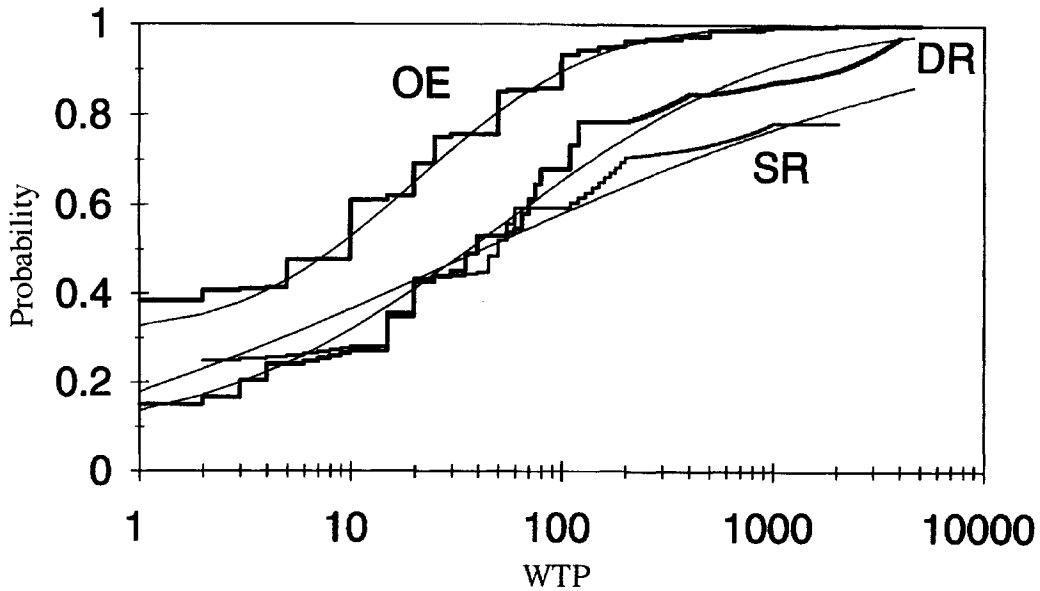


Figure 1. WTP cumulative distributions (light curves = parametric estimates)

response to discrete choice experiments on WTP is influenced by contextual effects, such as framing, anchoring, perceptions of fairness, self-image, and satisfaction with the interview process, then these effects may induce behavior in repeated bid referendum experiments that is not consistent with the maximization of a fixed utility function. I reject at the 1% level the null hypothesis of the same mean for the comparable experiments OE1 and SR1, and for the comparable experiments OE2 and SR2 (Wald statistic = 10.01, DF = 2, critical level = 9.21). The corresponding comparison of OE1 versus DR1, and OE2 versus DR2 also leads to rejection at the 1% level (Wald statistic = 25.16). The estimates of mean WTP given in table 3 show that there is an order-of-magnitude difference between the open-ended and referendum responses, with the latter giving far higher estimates of WTP. These differences are thus economically as well as statistically significant.

The hypothesis that the distribution of WTP is invariant with respect to elicitation format could be tested more stringently in the parametric setting by testing whether the vector of parameters (β, σ, π) is common between OE1 and SR1, etc. Since the weaker tests on comparability of means are rejected, this test would obviously also lead to a rejection. It is also possible to test comparability of estimated medians, which are less sensitive to the accuracy with which the parametric distributions approximate

the tails of the taste distributions. In each case, the referendum responses are four to five times the corresponding open-ended responses. The Wald statistic for OE1 versus SR1 and OE2 versus SR2 is 3.34, and the Wald statistic for OE1 versus DR1 and OE2 versus DR2 is 7.31. The second statistic is significant at the 5% level; the first statistic is not. Comparing the median contrasts with the mean contrasts, it is apparent that the elicitation format affects the entire WTP distribution, but the greatest differences are in the upper tails of the distributions.

I have also estimated nonparametrically the cumulative distribution functions (CDF) of WTP obtained by OE, SR, and DR response formats, and used these estimates to test whether the distributions are the same. The estimate for the open-ended responses is simply the empirical CDF. For the SR responses, I use the nonparametric maximum likelihood method of Cosslett, which estimates the CDF at each bid level by the sample frequency of "no" responses at that level, with bid levels grouped when necessary to give a monotone estimator. This method is also called the Pool-Adjacent-Violators-Algorithm (PAVA); see Robertson, Wright, and Dykstra. Finally, for the DR responses, let $F(B)$ denote the distribution function at bid B . Let $K_{NN}(B)$ denote the number of "no, no" responses observed at first bid B , and define $K_{NY}(B), K_{YV}(B), K_{YY}(B)$ similarly for the remaining three response patterns. Then, the

Table 5. Nonparametric Estimates of WTP Distribution

Treatment 1, Selway: OE Sample = 286, DR Sample = 426

WTP	Parametric			Empirical		
	OE	SR	DR	OE	SR	DR
1	0.328	0.179	0.137	0.385	0.250	0.152
2	0.355	0.230	0.172	0.409	0.250	0.167
4	0.408	0.287	0.224	0.416	0.261	0.243
5	0.432	0.306	0.245	0.479	0.275	0.243
10	0.530	0.367	0.321	0.612	0.282	0.273
20	0.650	0.431	0.413	0.692	0.436	0.429
30	0.723	0.469	0.472	0.759	0.443	0.453
40	0.772	0.496	0.516	0.759	0.450	0.532
50	0.807	0.517	0.550	0.853	0.522	0.532
60	0.834	0.534	0.578	0.857	0.595	0.548
80	0.872	0.561	0.621	0.860	0.595	0.681
100	0.897	0.581	0.654	0.934	0.595	0.681
120	0.916	0.598	0.680	0.944	0.651	0.787
200	0.954	0.643	0.750	0.965	0.707	0.787
400	0.983	0.701	0.830	0.976	0.733	0.848
500	0.988	0.718	0.852	0.990	0.766	0.848
1000	0.996	0.770	0.909	0.997	0.783	0.874
2000	0.999	0.815	0.948	1.000	0.783	0.901
Mean	44.7	3256.2	536.5	46.6	861.2	429.9
Median	8.3	41.7	36.1	10.0	46.9	35.9

Treatment 2, 57 Area: OE Sample = 298, DR Sample = 427

WTP	Parametric			Empirical		
	OE	SR	DR	OE	SR	DR
1	0.268	0.182	0.113	0.305	0.213	0.116
2	0.277	0.200	0.128	0.319	0.213	0.166
4	0.302	0.227	0.157	0.332	0.213	0.199
5	0.316	0.237	0.170	0.369	0.213	0.199
10	0.386	0.276	0.223	0.443	0.213	0.199
20	0.500	0.326	0.298	0.534	0.308	0.300
30	0.583	0.360	0.352	0.604	0.354	0.358
40	0.644	0.386	0.395	0.621	0.400	0.394
50	0.692	0.408	0.429	0.721	0.447	0.424
60	0.730	0.426	0.459	0.725	0.494	0.425
80	0.786	0.455	0.506	0.742	0.494	0.537
100	0.825	0.479	0.544	0.886	0.494	0.537
120	0.854	0.499	0.574	0.903	0.529	0.670
200	0.919	0.555	0.659	0.946	0.564	0.670
400	0.969	0.633	0.763	0.980	0.617	0.757
500	0.979	0.658	0.793	0.983	0.688	0.757
1000	0.994	0.731	0.871	0.993	0.723	0.832
2000	0.999	0.797	0.926	1.000	0.791	0.940
Mean	70.1	2880.1	717.5	70.2	923.4	463.9
Median	20.0	121.3	77.0	20.0	109.2	73.5

Goodness-of-Fit Tests	Selway	DF	57 Area	DF
Empirical OE vs SR	203.25	9	227.24	9
Empirical OE vs DR	183.76	18	225.64	18
Parametric vs Empirical OE	68.00	18	57.22	18
Parametric vs Empirical SR	136.89	9	74.00	9
Parametric vs Empirical DR	162.10	18	161.09	18

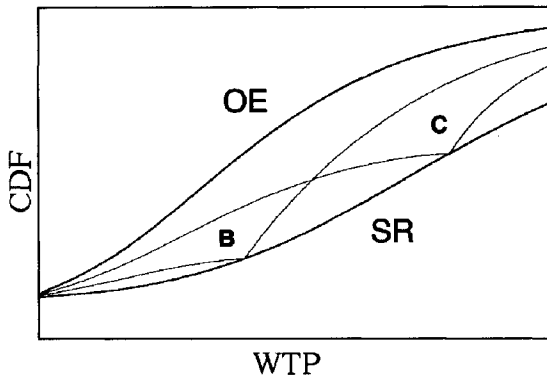


Figure 2. Referendum response pattern

contribution to log likelihood of the responses to first bid B is

$$\begin{aligned}
 &K_{NN}(B) \cdot \log[F(B/2)] \\
 &+ K_{NY}(B) \cdot \log[F(B) - F(B/2)] \\
 &+ K_{YN}(B) \cdot \log[F(2B) - F(B)] \\
 &+ K_{YY}(B) \cdot \log[1 - F(2B)].
 \end{aligned}$$

I maximize the sum of these contributions for the different first bids, treating F at each distinct bid as a separate parameter, subject to the restriction that F is monotone. The resulting estimates are given in table 5. The first panel contains the estimates for OE1, SR1, and DR1; for comparison, this panel also includes parametric estimates obtained from the mixed log normal models in table 3, with the conditioning on income removed using the empirical distribution of income for the subjects in all the Selway experiments. Linear interpolation is used to calculate SR1 at the second-bid levels in this table. The second panel contains the estimates for OE2, SR2, and DR2. The nonparametric estimates for the first panel are plotted in figure 1. The OE distribution is substantially shifted to the left relative to the referendum distributions, and the SR distribution has the thickest right tail. In the third panel of the table, I have calculated goodness-of-fit statistics for the hypotheses that the OE and SR distributions coincide, that the OE and DR distributions coincide, and that the parametric and nonparametric distributions coincide for each elicitation format. For the tests involving the SR distribution, I used the first bid levels to define categories, and for the remaining tests I used both first and second

bid levels to define categories. In all cases, the hypotheses are rejected at the 1% level. This testing procedure should be biased toward acceptance due to the imposition of monotonicity on the empirical distributions, and due to the estimation of parameters in the parametric distributions. The failure of the parametric distributions to match their nonparametric counterparts appears to arise partly from failure of the parametric models to capture tail behavior, but primarily from the “clustering” of responses around the values 5, 10, 20, 50, and 100. The nonparametric comparison of the CDFs indicates that although the parametric models fail to capture fully the behavior of the WTP distributions, the parametric analysis conclusion that different elicitation formats generate significantly different WTP distributions continues to hold.

A simple nonparametric test provides evidence on the validity of the DR method. An implication of utility maximization is that the probability of a “yes” to the lower of two bids and a “no” to the higher should be independent of the order in which the two bids are presented. For the first bids of \$10, \$20, \$100, or \$1,000 that yield a “yes,” followed by a “no” on the doubled bid, and the first bids of \$20, \$40, \$200, or \$2,000 that yield a “no,” followed by a “yes” on the halved bid, one obtains the same bid pairs. The test statistic is a sum over the DR experiments of the expressions $P'V^{-1}P$, where

$$P' = [p_{10}^{YN} - p_{20}^{NY}, p_{20}^{YN} - p_{40}^{NY}, p_{100}^{YN} - p_{200}^{NY}, p_{1000}^{YN} - p_{2000}^{NY}],$$

V is the matrix

$$\begin{bmatrix}
 v_{10}^{YN} + v_{20}^{NY} & c_{20} & 0 & 0 \\
 c_{20} & v_{20}^{YN} + v_{40}^{NY} & 0 & 0 \\
 0 & 0 & v_{100}^{YN} + v_{200}^{NY} & 0 \\
 0 & 0 & 0 & v_{1000}^{YN} + v_{2000}^{NY}
 \end{bmatrix}$$

where p_k^{YN} is the sample frequency of a “yes” followed by a “no” at initial bid k , $v_k^{YN} = p_k^{YN}(1 - p_k^{YN})/n_k$ is an estimate of the variance of the associated sample frequency with n_k observations, $c_{20} = p_{20}^{YN} p_{20}^{NY}/n_{20}$ is an estimate of the sample covariance of the differences $p_{10}^{YN} - p_{20}^{NY}$ and $p_{20}^{YN} - p_{40}^{NY}$, etc. I reject at the 1% level the hypothesis that first and second responses in the double referendum experiment are drawn

from the same distribution (statistic = 51.2, DF = 16, critical level = 32.0).

For the paired bid levels in this test, the average of the p_k^{NY} over bids and experiments is 0.204, while the average of the p_k^{OE} is 0.088. Figure 2 provides a graphical characterization of this pattern and the large shift upward in referendum responses compared to OE responses. In this figure, there is a cumulative distribution of open-ended WTP response, denoted OE, and a distribution of first-bid referendum response, denoted SR. A first bid B provides an anchor that modifies the distribution of second responses to BB. A higher first bid C modifies the distribution of second responses to CC. Then, the probability of "yn" is larger than the probability of "ny".

The CV response patterns summarized in figure 2 have several possible interpretations. One is that OE represents "true" preferences, and SR responses are distorted by the "anchoring" effect of the first bid, which respondents may weigh as a cue to the actual cost of preservation, or to socially approved response levels. In this interpretation, a second referendum bid far away from the first "delegitimizes" the anchor, and second responses more closely resemble OE responses, while a second referendum bid near the first leads to relatively little modification of behavior. A related interpretation is that there is a psychological cost to giving a "no" response when a "yes" is recognized as the polite or moral alternative. This induces the difference between OE and SR. Once this cost is either avoided by an initial "yes," or incurred by an initial "no," it is much easier to say "no" to subsequent questions. An alternative interpretation is that SR represents "true" preferences, and subsequent referendum responses are distorted due to the anchoring effect of the first bid. A problem with this interpretation is that a pure anchoring effect would be expected to sharpen discrimination around the first bid, whereas the observed pattern skews the second response density downward.

Conclusions

Summarizing the results from these experiments, I find that CV responses are quite sensitive to psychometric context effects, particularly anchoring to cues provided by the experimenter, and are affected by outliers that make statistical analysis difficult. There is both

parametric and nonparametric evidence that response behavior is different between first response and second response in double referendum surveys. Thus, the "starting point bias" that led CV researchers to abandon repetitive bidding games is already a damaging effect in second response, and the double referendum elicitation format is internally inconsistent. I find in parametric tests, and confirm in nonparametric comparisons, that the open-ended format yields a WTP distribution that is significantly different from either the single-referendum or double-referendum formats. The current experiments cannot identify a "reliable" elicitation format, or rule out the possibility that neither open-ended or single-referendum responses are good approximations to true preferences. One cannot even presume that if a true WTP distribution exists, it is bracketed by single referendum and open-ended responses.

The experiments display patterns that are more easily explained by "constructed" preferences than by rational individualistic stationary preferences. While the magnitude of the effects observed in this experiment may depend on its particular design, including the survey method and the clarity and cogency of the resource issue, they are nevertheless typical of psychometric distortions that pervade market research and psychological experiments.¹³ By incorporating in their CV studies an experimental design like the one in this paper, practitioners could test whether they are in fact successful in avoiding these problems for commodities with existence value.

Obviously, if CV is to be credible for the estimation of existence values, it should at least work reliably for market commodities, or for issues of fact. An immediate research priority should be to conduct experiments that identify the biases that arise from CVM in these applications, and test whether there are modifications to CV methodology that can avoid these biases in the survey instrument or correct them in the data analysis. Possible avenues to investigate include giving the consumer comprehensive exposure to the resource and to the impacts of alternative policies, perhaps over an extended period of time; training the consumer to think in terms of value and price of analogous

¹³ A battery of critical studies of the CV method are summarized by Diamond and Hausman. Other studies that raise substantive questions about the features of the method are Bishop and Heberlein; Kahneman and Knetsch; Kealy, Montgomery, and Davidio; Loomis; McDaniels; and McClelland et al.

public and private goods; and creating a transaction where money changes hands, so that responses have a cost. It may also help to collect detailed information on components and rationalizations of value, using conjoint analysis methods from market research, attitude inventories, verbal protocols, and debriefing data, so that sources of bias can be identified and isolated. If evidence accumulates that CV measurement of existence value is encountering fundamental failures of preference theory, in which preferences are "constructed" for each situation and are highly sensitive to position and context, then more sophisticated survey methods that strip away one level of cognitive distortions will simply reveal, or induce, new distortions. In this case, it will be necessary to turn to mechanisms other than polling consumers in order to achieve a reliable basis for valuing these resources.

[Received October 1992;
final revision received February 1994.]

References

- Bishop, R., and T. Heberlein. "Measuring the Values of Extra-Market Goods: Are Indirect Measures Biased?" *Amer. J. Agr. Econ.* 61(December 1979):926-30.
- Bowker, J., and J. Stoll. "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." *Amer. J. Agr. Econ.* 70(May 1988):372-81.
- Cameron, T., and D. Huppert. "Referendum Contingent Valuation Estimates: Sensitivity to the Assignment of Offered Values." *J. Amer. Statist. Assoc.* 86(December 1991):910-18.
- Cameron, T., and J. Englin. "Cost-Benefit Analysis for Non-Market Resources Under Individual Uncertainty." U.C.L.A. Working Paper, 1992.
- Cameron, T., and M. James. "Efficient Estimation Methods for Use with 'Closed Ended' Contingent Valuation Survey Data." *Rev. Econ. and Statist.* 69(May 1987):269-76.
- Carson, R. "Constructed Markets." J. Braden and C. Kolstad, eds. *Measuring the Demand for Environmental Quality*, pp. 121-62. Amsterdam: Elsevier, 1991.
- Cosslett, S. "Distribution-Free Maximum Likelihood Estimation of the Binary Choice Model." *Econometrica* 51(May 1983):765-82.
- Cummings, R., D. Brookshire, and W. Schultz. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Savage MD: Rowman & Littlefield, 1986.
- Diamond, P., and J. Hausman. "On Contingent Valuation Measurement of Nonuse Values." J. Hausman, ed. *Contingent Valuation: A Critical Assessment*, pp. 3-38. Amsterdam: Elsevier, 1993.
- Diamond, P., J. Hausman, G. Leonard, and M. Denning. "Does Contingent Valuation Measure Preferences? Experimental Evidence." J. Hausman, ed. *Contingent Valuation: A Critical Assessment*, pp. 41-85. Amsterdam: Elsevier, 1993.
- Duffield, J., and D. Patterson. "Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation." *Land Econ.* 67(May 1991):225-39.
- Hanemann, M. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *Amer. J. Agr. Econ.* 66(August 1984):332-41.
- Hanemann, M., J. Loomis, and B. Kanninen. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *Amer. J. Agr. Econ.* 73(November 1991):1255-63.
- Hausman, J., G. Leonard, and D. McFadden. "The Value of Recreational Opportunities Lost Due to the Exxon Valdez Oil Spill." J. Hausman, ed. *Contingent Valuation: A Critical Assessment*, pp. 341-85. Amsterdam: Elsevier, 1993.
- Kahneman, D., and J. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *J. Environ. Econ. and Manage.* 22(January 1992):57-70.
- Kealy, M., M. Montgomery, and J. Davidio. "The Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?" *J. Environ. Econ. and Manage.* 19(November 1990):244-63.
- Loomis, J. "Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques." *J. Environ. Econ. and Manage.* 18(January 1990):78-85.
- Loomis, J. "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models." *Land Econ.* 68(May 1992):211-24.
- McClelland, G., W. Schultze, J. Lazo, D. Waldman, J. Doyle, S. Elliott, and J. Irwin. *Methods for Measuring Non-Use Values: A Contingent Valuation Study of Groundwater Cleanup*. University of Colorado Working Paper, 1992.
- McConnell, K. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *J. Environ. Econ. and Manage.* 18(January 1990):19-34.
- McDaniels, T. "Reference Points, Loss Aversion,

- and Contingent Values for Auto Safety." *J. Risk and Uncertainty* 5(May 1992):187-200.
- McFadden, D. "Estimation of Social Value from Willingness-to-Pay Data." J. Moore, ed. *Essays in Honor of John Chipman*, forthcoming (1994).
- _____. "The Choice Theory Approach to Market Research." *Mktg. Sci.* 5(Fall 1986):275-97.
- McFadden, D., and G. Leonard. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." J. Hausman, ed. *Contingent Valuation: A Critical Assessment*, pp. 165-208. Amsterdam: Elsevier, 1993.
- Park, T., J. Loomis, and M. Creel. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies." *Land Econ.* 67(February 1991):64-73.
- Rao, C.R. *Linear Statistical Models*. New York: Wiley, 1973.
- Robertson, T., F. Wright, R. Dykstra. *Order Restricted Statistical Inference*. New York: Wiley, 1988.