TIM PHIPPS
RESOURCE ECONOMICS

WESTERN REGIONAL RESEARCH PUBLICATION

W-133
Benefits and Costs Transfer in Natural Resource Planning

Eighth Interim Report

October 1995

Compiled by
Douglas Larson

Department of Agricultural Economics
University of California, Davis
Davis, CA 95616

CONTENTS

OBJECTIVE 2: DEVELOPING PROTOCOLS FOR TRANSFERRING VALUE ESTIMATES

TOWARDS AN IMPROVED UNDERSTANDING OF VALUATION METHODS

Contingent Valuation

Random Utility and Resource Valuation Modelling

INTRODUCTION

This volume contains the proceedings of the W-133 Regional Research Project Annual Technical Meeting held in Pacific Grove, California, in March 1995. The purposes of this Western Regional Research project include the economic valuation of environmental amenities and natural resources; the use of these values to inform public policy and decisionmaking; and the study of how values from pre-existing studies may be credibly "transferred" from one resource or geographic area to new ones. Researchers from more than 25 land grant institutions around the country are formally involved in the W-133 project through their campus Agricultural Experiment Stations, and the project attracts many more participants from federal and state agencies as well as researchers from other institutions interested in valuation questions. This participation occurs both through conducting cooperative research efforts addressed to one or more of the objectives or resource areas of W-133, and through attending and presenting papers at the annual Technical Meeting. The interaction and cooperation among a broad spectrum of resource managers and researchers is one of the unique strengths of W-133.

The specific objectives of W-133 are to (1) provide site-specific use and nonuse values of natural resources for public policy analyses; and (2) to develop protocols for transferring value estimates to unstudied areas. Research conducted by W-133 participants to meet these objectives is targeted at four resource areas: water-based recreation, groundwater quality, wetlands, and recreational fisheries. In addition to the many case studies of amenity values and benefits transfer exercises which investigators conduct, many fundamental research methodology questions are encountered in the area of nonmarket valuation. Making progress toward resolving these questions is essential to increasing confidence in the empirical value estimates, so many investigators and cooperators also present and discuss research aimed at these methodological questions.

This volume is organized around the objectives and resource areas of the W-133 project. The first three sections address objective 1 of the project, site-specific use and nonuse values in the resource areas of water-based recreation, groundwater quality and rural amenities, and recreational fisheries and hunting studies. The next section addresses the objective of benefits transfer. The final two sections address some of the fundamental methodological issues that run throughout all efforts to provide convincing resource values, both in the areas of contingent valuation and in the areas of random utility and resource valuation modelling.

Any classification scheme is, to an extent, arbitrary. Many of the papers in this volume cross category lines as defined above and provide insight into more than one area. In particular, the papers addressing site-specific values often must also address methodological issues. Conversely, the papers which highlight methodology are also often based on empirical studies that provide site-specific or amenity-specific values. However one chooses to classify them, the papers in this volume amply demonstrate the rich variety and high quality of research into the important area of amenity valuation which the W-133 project makes possible.

In closing, I would like to acknowledge the capable assistance of Dan Lew in preparing this volume.

Douglas M. Larson
University of California, Davis
September 1995

SYSTEMS OF TRAVEL COST MODELS OF RECREATION DEMAND

by

J.S. Shonkwiler
Professor
Department of Agricultural Economics
University of Nevada
Reno, Nevada 89557
(702) 784-1341

Comments by Jeffrey Englin and Douglass Shaw are gratefully acknowledged. Discussions with and research materials provided by Jeffrey LaFrance were particularly helpful. Research supported by the Nevada Agricultural Experiment Station via Western Regional Research Project W-133.

SYSTEMS OF TRAVEL COST MODELS OF RECREATION DEMAND

Recreation demand modeling is an important element of natural resource planning. Behavioral responses and valuations of recreationists are often used as components of benefit-cost analysis or environmental impact assessment. The travel cost model which defines a demand function for recreation sites has been employed by economists since the early 1960's (Smith, 1989). Yet a number of theoretical and empirical problems encompass the travel cost model. These include issues involving the count data structure of the dependent demand quantity, assumptions regarding the structure of the demand decision relative to corner solutions and hurdles to consumption, and the treatment of multiple sites in the empirical specification.

Count data travel cost models have become increasingly more common (Creel and Loomis, 1990; Hellerstein, 1991; Englin and Shonkwiler, 1995) as economists have recognized that travel cost studies based on participant information regarding visits to recreation sites are subject to the fact that each respondent will report a discrete number of trips. If the latent demand for travel to a recreation site is considered to lie in the interval $(-\infty, \infty)$, then since observed trip demand cannot be negative, it is censored at zero and failure to account for censoring leads to biased estimators. The application of count data estimators to the travel cost model thus is a logical extension to accommodate the particular properties of trip data. In view of the recent work of Hellerstein and Mendelsohn, who provide theoretical foundations for linking the empirical count estimator to the individual consumer's underlying optimization problem, it is clear how to perform welfare analysis on the basis of a single equation count demand model for trips.

Yet a single, independent recreational site rarely exists. If similar recreational experiences can be obtained at sources near the recreation site of interest there may be a high degree of substitutability among such sites. Although most studies to date have assumed independence in order to estimate demand, researchers recognize the probably important interdependencies of demands for sites due to the pioneering work of Burt and Brewer. Subsequent studies by Cicchetti et al. and Sellar et al. have provided additional evidence to justify a systems approach.

Unfortunately, analyses of household (or individual) demands for recreation have not accounted for the discrete, zero and positive integer, characterization of trip data. The single published exception is the recent study by Ozuna and Gomez. Ozuna and Gomez, however, adopt a highly restrictive econometric model which cannot accommodate over-dispersion or negative covariances between equations, and they fail to recognize that welfare analysis in a systems context is altered when conditional expected demands have an exponential form.

With regard to corner solutions and hurdles to consumption, Pudney has pointed out that individuals may make non-marginal changes by switching from one behavioral regime to another. The two types of responses involved are at the intensive margin where consumers of the good are motivated to consume marginally more and at the extensive margin where people may either enter or leave the market entirely. Individuals with zero consumption may be at a corner solution such that a price reduction (income increase) may lead to non-zero levels of consumption. Alternatively, zero consumption may represent behavior which is robust to changing economic variables, reflecting instead a choice set influenced by endowments or physical capabilities. While econometric models developed by Mullahy may provide insight into the treatment of zero consumption as the consequence of a corner solution to the conventional utility maximization problem, they are unsatisfactory in terms of their ability to distinguish situations where desired consumption may be positive but observed consumption is recorded as zero.

This paper attempts to synthesize the elements necessary to appropriately treat multiple site travel cost models of recreation demand when the decision variables are measured as trip counts. A flexible, multivariate count data probability model is adopted and modified to account for the excess zero consumption problem. Because this model generates conditional demands with exponential form, a proper incomplete demand structure (LaFrance and Hanemann) will be imposed to insure that exact welfare analysis can be performed. The proposed techniques will then be applied to a stylized empirical model of angler demand for visits to three lakes in northwestern Nevada.

The Multivariate Poisson-Log Normal Distribution

While there are a number of multivariate discrete distributions, Aitchison and Ho argue that only the multivariate Poisson-log normal (MPLN) distribution can both reproduce an arbitrary correlation structure and account for over-dispersion. This distribution arises from mixing the location parameters of independent Poisson processes with a multivariate log normal distribution. Whereas the univariate Poisson-log normal distribution is well known (Johnson and Kotz), the multivariate form was apparently not proposed until Aitchison and Ho's 1989 paper.

To motivate the development of the MPLN econometric model, first consider its univariate form. Let q_{ni} denote the n^{th} households observed demand for the i^{th} recreation site. Suppressing the site index, the Poisson probability mass function for q_n is

(1) $$f(q_n|\theta) = \theta^{q_n} e^{-\theta} / q_n! \quad \text{and} \quad E(q_n) = \theta$$

It is well known that, if the Poisson parameter θ is mixed with a Gamma distribution and integrated out, the negative binomial distribution results. When the Poisson parameter is mixed with the log normal distribution, the following probability mass function results.

(2) $$g(q_n(\theta)|\mu,\sigma^2) = \int_0^\infty \frac{\theta^{q_n} e^{-\theta} \exp\left(-\tfrac{1}{2}(\ln\theta - \mu)^2/\sigma^2\right)}{q_n!\sqrt{2\Pi}\,\theta\sigma} d\theta$$

The log normal is chosen as a mixing distribution for several reasons: i) it is defined over positive values only; ii) it can represent highly skewed data; and iii) it easily generalizes to a multivariate form. Property (i) is important because the Poisson parameter is strictly greater than zero. Property (ii) is attractive because frequently a few households or individuals will have demands many standard deviations greater than the average. Property (iii) permits treatment of related sites.

Note than in (2) θ is a nuisance value which must be integrated out. Unfortunately, no closed form exists for this integral so that to obtain maximum likelihood parameter estimates, the expression

(3) $$\sum_{n=1}^{N} \ln g(q_n(\theta)|\mu,\sigma^2)$$

must be maximized with respect to μ and σ^2. In terms of moments

$$
(4)\qquad \begin{aligned} &E(q_n) = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right) \text{ and} \\ &\operatorname{Var}(q_n) = E(q_n) + \left(\exp\left(\sigma^2\right) - 1\right)\left(E(q_n)\right)^2 \end{aligned}
$$

so it is apparent that if $\sigma^2 = 0$ then a Poisson model with $\theta = \exp(\mu)$ could reproduce the first two moments of the univariate Poisson-log normal distribution.

To develop the multivariate form of (2), we suppress the observational subscript n and let θ and μ denote G element vectors corresponding to G related sites. Σ denotes the GxG covariance matrix with element $\sigma_{ij}, i = 1,2,\ldots G$ and $j = 1,2\ldots G$. Then the MPLN joint probability mass function becomes

$$
(5)\qquad \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{G} \frac{f(q_i|\theta_i)\exp\left\{-\tfrac{1}{2}(\ln\theta - \mu)'\,\Sigma^{-1}(\ln\theta - \mu)\right\}}{(2\Pi)^{\frac{1}{2}G}(\theta_1\theta_2\cdots\theta_G)|\Sigma|^{\frac{1}{2}}}\,d\theta = P_G\left(q(\theta)|\mu,\Sigma\right)
$$

for the n^{th} observation when $f(.)$ is defined in (1). The implied covariances have the form

$$
(6)\qquad \begin{aligned} Cov(q_i, q_j) &= E(q_i)E(q_j)\left(\exp(\sigma_{ij}) - 1\right) \quad \text{given} \\ E(q_i) &= \exp\left(\mu_i + \tfrac{1}{2}\sigma_{ii}\right) \end{aligned}
$$

thus the conditions $\sigma_{ij} < 0$, $\sigma_{ij} = 0$, and $\sigma_{ij} > 0$ establish negative, zero and positive covariances respectively between the observed sites.

The price of the flexibility of the MPLN probability density is the G dimensional integrals which must be evaluated for each of $n = 1,2,\ldots N$ observations in order to evaluate the likelihood for the sample. Furthermore, since the derivatives of the likelihood with respect to μ and Σ cannot be factored out of (5), the multidimensional integral must be evaluated for each derivative (either numerical or analytical) computed. Computational approaches will be discussed more extensively in the empirical section.

While the likelihood function represented by equation (5) may not be simplified, quasi-maximum likelihood estimation methods can be employed to reduce the dimensionality of the integral by specifying a sequence of, say, bivariate likelihood functions. For example a three equation system could be approximated by the product of three bivariate likelihoods

$$(7)\qquad \prod_{i=1}^{j-1}\prod_{j=2}^{1/2G(G-1)}\int_0^\infty\!\!\int_0^\infty \frac{\theta_i^{q_i}\,\theta_j^{q_j}\,e^{-\theta_i-\theta_j}\exp\left\{\dfrac{-1}{2-2\rho_{ij}^2}\left[\left(\dfrac{\ln\theta_i-\mu_i}{\sigma_i}\right)^2+\left(\dfrac{\ln\theta_j-\mu_j}{\sigma_j}\right)^2 x\right.\right.}{2\Pi\left(1-\rho_{ij}^2\right)^{1/2}\sigma_i\sigma_j q_i!q_j!\theta_i\theta_j}$$

$$\left.\left.-2\rho_{ij}\left(\ln\theta_i-\mu_i\right)\left(\ln\theta_j-\mu_j\right)/\sigma_i\sigma_j\right]\right\}d\theta_i d\theta_j$$

In this manner, evaluation of a triple integral is reduced to evaluation of three double integrals. Estimated standard errors of the parameters will need to be adjusted with the formula in White since the parameters which maximize (7) are quasi-maximum likelihood estimators.

Hurdles/Excess Zeros in Count Data Models

It is known that non-consumption of a particular site, or possibly a group of sites, is a commonly observed phenomenon having both economic and statistical implications. Let q_n denote the number of visits the n^{th} (potential) user of a single specific recreation site has made over a season. Define two variables that influence the individual's decisions as μ_n and ω, with the former conditioned by mainly economic variables (prices, income) and the latter independent of economic variables. This decomposition of explanatory variables is in the spirit of Pudney who suggests separating economic variables from personal characteristics which shape tastes. Let D_n represent the latent decision to consume such that consumption is zero if $D_n \leq 0$. Specify

$$(8)\qquad \Pr(D_n \leq 0) = \omega$$

where ω is an unknown parameter.

If consumption is positive, then it is assumed that observed consumption equals desired consumption

$$(9)\qquad q_n = q_n^* \quad \text{with}$$
$$E(q_n^*) = \mu_n$$

Mullahy has proposed a single hurdle model of consumption behavior that Shonkwiler has criticized because the decision to consume is independent of the level of consumption (vid Morey et al. regarding the discussion that non-participation represents only one of many types of boundary solutions). Instead, Shonkwiler formulated a double hurdle model of consumption behavior originally suggested by Blundell and Meghir which provides the regimes

$$\text{(10)} \quad \begin{aligned} &\Pr(q_n = 0) = \Pr(q_n^* = 0) + \Pr(q_n^* > 0)\Pr(D_n \leq 0) \\ &\text{and} \\ &\Pr(q_n > 0) = \Pr(q_n^* > 0) - \Pr(q_n^* > 0)\Pr(D_n \leq 0) \end{aligned}$$

No consumption will be observed if desired consumption is non-positive, or, if desired consumption is positive, an additional hurdle $(D \leq 0)$ may prevent consumption. The probability of a positive observation is determined by whether latent demand is positive and whether the binary impediment to consumption is effective.

As shown by Shonkwiler this model incorporates two mechanisms for generating zeros, one intended to represent a fundamental non-economic decision and the other representing an ordinary corner solution (Pudney). Additionally, this hurdle mechanism corresponds to the excess zero probability model of Johnson and Kotz. Recognizing that (10) can be rewritten

$$\text{(11)} \quad \begin{aligned} &\Pr(q_n = 0) = \Pr(D_n \leq 0) + \Pr(D > 0)\Pr(q_n^* = 0) \\ &\Pr(q_n > 0) = \Pr(D_n > 0)\Pr(q_n^* > 0) \end{aligned}$$

we obtain the zero modified probability mass functions of Johnson and Kotz. Namely

$$\text{(12)} \quad \begin{aligned} &\Pr(q = 0) = \omega + (1-\omega)P_0 \\ &\Pr(q = j) = (1-\omega)P_j \qquad j = 1,2,\ldots \end{aligned}$$

where the original distribution has probability mass function $P_j, j = 0,1,2\ldots$ These zero modified distributions can be extended to multivariate discrete distributions in a similar fashion

$$\text{(13)} \quad \begin{aligned} &\Pr(q_1 = q_2 = \ldots q_G = 0) = \omega + (1-\omega)P_G(0,0,\ldots 0) \\ &\Pr(q) = (1-\omega)P_G(q) \qquad q_1 = q_2 \ldots = q_G \neq 0 \end{aligned}$$

Johnson and Kotz point out that all uncentered moments of the modified distribution differ from the corresponding moments of the original distribution by the factor $(1-\omega)$.

Incomplete Demand Systems

Specification of a system of demand equations naturally leads to the implications of consumer choice theory for assessing the structure imposed. As LaFrance has pointed out, three practical approaches can be considered for the demand system specification. First, broad aggregates of all goods available to the consumer can be used to reflect all choices in the consumption set. Second, separability can be imposed so that conditional demand equations involving a subset of commodities can be estimated. Third, an incomplete system of demand equations can be specified. Obviously, the first approach is unsatisfactory because interest is focused on individual commodities. The second approach suffers from i) uncertainty as to the true nature of separability, ii) not identifying the overall utility function but only a subutility function, and iii) the interdependence between quantities demanded and group expenditure. This latter condition is exacerbated when many households have zero demands and consequently zero groupwise expenditure. Thus, substantial simultaneous equations bias would likely be encountered.

The incomplete demand system specification is an attractive alternative only if the preference structure it identifies is consistent with rational models of consumer behavior. Incomplete demand models that can be related to an underlying utility maximization subject to a linear budget constraint can be used to conduct proper welfare analysis (LaFrance and Hanemann, 1989). The incomplete demand structures that are consistent with such maximizing behavior were first presented in LaFrance and Hanemann (1984). For linear expected demands, the restrictions typically imposed are zero (or essentially zero) income effects and a symmetric negative definite cross price matrix. Burt and Brewer as well as Seller et al. imposed cross equation symmetry of the price coefficients. Hence both studies imposed restrictions generally consistent with those suggested by a linear incomplete demand system. However, because both studies modeled discrete household demand data with linear models, their welfare calculations are compromised by their linear estimators.

Ozuna and Gomez tested symmetric cross equation price effects stating that this is a necessary condition for path independence of the line integral used to compute the welfare effects of price changes. But Ozuna and Gomez specified their expected demands with an exponential form

$$(14) \qquad E(q_i) = \exp\left(\alpha_i + \sum_j \beta_{ij} p_j + \gamma_i y\right) \qquad i = 1,2,\ldots G$$

where y represents household income and the observational subscript has been suppressed. Symmetry conditions for this model are

$$(15) \qquad \beta_{ij} q_i + \gamma_i q_i q_j = \beta_{ji} q_j + \gamma_j q_i q_j$$

so, irrespective of whether the γ_i are essentially zero, symmetry will generally not hold unless $\beta_{ij} = \beta_{ji} = 0$.

Conditional expectations from the zero modified MPLN probability model have the form

$$(16) \qquad E(q_i) = (1-\omega)\exp\left(\alpha_i + \beta_{ii} p_i + \gamma y + \tfrac{1}{2}\sigma_i^2\right).$$

Here the restrictions consistent with an incomplete exponential demand system (LaFrance and Hanemann, 1984) have been imposed, viz. $\gamma_i = \gamma$ and $\beta_{ij} = 0,\ i \neq j$ The demand shifters, or a_i in the terminology of LaFrance, are defined by

$$(17) \qquad a_i = (1-\omega)\exp\left(\alpha_i + \tfrac{1}{2}\sigma_i^2\right)$$

and a final restriction requires

$$(18) \qquad a_i = \frac{\beta_{ii}}{\beta_{11}} a_1 \qquad \forall\, i$$

so that in terms of each $\alpha_i (i \neq 1)$ we have

$$(19) \qquad \alpha_i = \ln(\beta_{ii}/\beta_{11}) + \alpha_1 + \tfrac{1}{2}\left(\sigma_1^2 - \sigma_i^2\right)$$

Thus the exponential incomplete demand system has G free own-price parameters, one income coefficient and one intercept parameter plus the $\tfrac{1}{2}G(G+1)$ unique elements of Σ.

Although the restrictions imposed on the incomplete demand system appear severe, there is some latitude for alternative specifications which still satisfy the utility theoretic requirements (LaFrance). One alternative is to explicitly introduce complementarity in the demand system by imposing restrictions which do not require zero cross price coefficients. However, the specification that is adopted herein employs a maintained hypothesis that the specific recreation sites considered are substitutes and the degree of substitution, s_{ij}, can be expressed by recognizing that the Slutsky equation has the form

$$s_{ij} = q_j \frac{\partial q_i}{\partial y} = \gamma q_i q_j \geq 0 \quad \text{if} \quad \gamma > 0$$

under the restrictions imposed and the statistical model specified.

Empirical Analysis

The data to be analyzed were collected in 1988 through a cooperative effort by the Nevada Department of Wildlife and the Department of Agricultural Economics, University of Nevada. Residents holding valid Nevada fishing licenses were surveyed by mail. Respondents were asked to supply information to questions regarding household characteristics (number of family members, age and gender of family members, education levels), household income, and expenditures on equipment, guide and taxidermy services. Respondents were also asked to provide single and multiple site trip information regarding visits to a number of Nevada lakes and reservoirs.

For the purposes of this stylized study, visits to Topaz Lake, Walker Lake and Pyramid Lake by northern Nevada anglers are analyzed in order to infer their use values to area anglers. All lakes are located in northwestern Nevada and sample statistics are reported in Table 1. A total of 532 observations comprised the data set once missing values and anglers reporting more than 30 trips were dropped. Note that anglers who visit a given site tend to make multiple trips, whereas a substantial percentage of angers make no visits (Table 1). Explanatory variables used are y_n, the n^{th} household's income in \$10,000, and p_{ni}, the n^{th} household's round-trip distance to the i^{th} recreation site multiplied by \$.25, to approximately represent travel cost. For some indication of how accurately this measure represents the true, unobserved travel price, see the recent study by Englin and Shonkwiler.

To provide a basis of comparison, univariate Poisson-log normal and negative binomial models are estimated for each of the three lakes. Both distributions were modified to account for excess zeros and this probability is represented by the ω parameter. For the Poisson-log normal model

$$\mu_{ni} = \alpha_i + \beta_i p_{ni} + \gamma_i y_{ni}$$

and for the negative binomial model the location parameter is specified as

$$\theta_{ni} = \exp(\alpha_i + \beta_i p_{ni} + \gamma_i y_{ni})$$

and the dispersion parameter is denoted ϕ. Results are presented in Table 2. Notice that the Poisson-log normal model produces comparable or better fits as judged by the log likelihood values.

Full information maximum likelihood estimation of the three equation system requires evaluation of triple integrals at each data point. Two approaches were used to calculate the multidimensional integrals. The intquad3 procedure in GAUSS with order of integration set at 32 was used to evaluate equation (5) and multivariate Hermite integration was used to evaluate a transformation of (5) suggested by Aitchison and Ho. The first approach appeared to assign too much weight to observations of zero demand and the second approach appeared to overweight the probabilities associated with large demands. Thus, the estimation results from both models were pooled by taking geometric averages of respective parameters.

Results for the pooled FIML estimates under the restrictions generated by the incomplete demand system structure are presented in Table 3. Parameters correspond to those appearing in equation (16). Also, the multivariate distribution is modified to account for excess zeros as per equation (13). These findings establish significant negative correlations between demands for trips to Topaz and Walker Lakes as well as Topaz and Pyramid Lakes. Marginally significant positive correlation is found between Walker and Pyramid Lakes. The FIML system specification also permits proper imposition of the cross-equation parameter restrictions required by the incomplete demand structure. Notice that for the full system, we find that the probability of an impediment to consumption (as opposed to a corner solution) is about 38 percent.

Table 3a presents the quasi-maximum likelihood parameter estimates for the same three equation system but using likelihoods defined by equation (7). This method generates correlations which appear to be uniformly smaller than those of the FIML system. On the other hand, the own price and income coefficients differ by no more than six percent. Given that computational time is reduced about one order of magnitude using the quasi-maximum likelihood approach, this technique appears promising particularly for higher dimensioned problems.

Of course the primary purpose of the incomplete demand specification is to permit the calculation of appropriate welfare measures. Table 4 presents these measures for the average angler in our sample. Because of the small magnitude of the income coefficients, the results show equivalent and compensating variation tightly bracketing consumer's surplus. One interesting application of these results is to assess the use value of Walker Lake to northern Nevada anglers. This is a reasonable consideration because the lake is slowly dying because of increasing salinity due to extended drought and over-commitment of Walker River water to upstream interests. No river water has flowed into the lake in the last eight years. Assuming there are 40,000 area anglers who put a use value of $12.03 on the lake per season, we find using a five percent discount rate, a $10,336,000 present value (1988 dollars) for Walker Lake.

Conclusions

The Poisson-log normal probability model appears to be a flexible alternative to the negative binomial probability mass function. Because the Poisson-log normal probability model can be generalized to the multivariate case and represent any arbitrary pattern of correlation among equations, the computational burden incurred in its implementation seems justified. Further, it can be specified to accommodate a double hurdle model of behavior which is demonstrated to have substantial empirical importance. Finally, the incorporation of the incomplete demand system restrictions for exponential specifications of expected demand are seen to provide two critical benefits. These restrictions reduce the number of free parameters and permit exact welfare analysis.

References

Aitchison, J. and C.H. Ho. "The Multivariate Poisson-Log Normal Distribution." Biometrika, 76(1989): 643-653.

Blundell, R. and C. Meghir. "Bivariate Alternatives to the Tobit Model." Journal of Econometrics, 34(1987): 179-200.

Burt, O.R. and D. Brewer. "Estimation of Net Social Benefits from Outdoor Recreation." Econometrica, 39(1971): 813-827.

Cicchetti, C.J., A.C. Fisher and V.K. Smith. "An Econometric Evaluation of a Generalized Consumer Surplus Measure: The Mineral King Controversy." Econometrica, 44(1976): 356-360.

Creel, M. and J. Loomis. "Theoretical and Empirical Advantages of Truncated Count Data Estimators." American Journal of Agricultural Economics, 72(1990): 434-441.

Englin, J. and J.S. Shonkwiler. "Estimating Social Welfare Using Count Data Models." Review of Economics and Statistics, (forthcoming).

Englin, J. and J.S. Shonkwiler. "Modeling Recreation Demand in the Presence of Unobservable Travel Costs: Toward a Travel Price Model." Journal of Environmental Economics and Management, forthcoming.

Hellerstein, D.M. "Using Count Data Models in Travel Cost Analysis with Aggregate Data." American Journal of Agricultural Economics, 73(1991): 860-866.

Hellerstein, D.M. and R. Mendelsohn. "A Theoretical Foundation for Count Data Models." American Journal of Agricultural Economics, 75(1993): 604-611.

Johnson, N. and S. Kotz. Discrete Distributions. 2nd edition. New York: John Wiley, 1992.

LaFrance, J.T. and W.M. Hanemann. "On the Integration of Some Common Demand Systems." Staff Papers in Economics 83-10, Montana State University, October, 1984 (rev.)

LaFrance, J.T. and W.M. Hanemann. "The Dual Structure of Incomplete Demand Systems." American Journal of Agricultural Economics, 71(1989): 262-274.

LaFrance, J.T. "Incomplete Demand Systems and Semilogarithmic Demand Models." Australian Journal of Agricultural Economics, 34(1990): 118-131.

Morey, E.R., D. Waldman, D. Assane and D. Shaw. "Searching for a Model of Multiple-site Recreation Demand that Admits Interior and Boundary Solutions." American Journal of Agricultural Economics, forthcoming.

Mullahy, J. "Specification and Testing of Some Modified Count Data Models." Journal of Econometrics, 33(1986): 341-365.

Ozuna, T. and I.A. Gomez. "Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach." Review of Economics and Statistics, 76(1994): 356-360.

Pudney, S. Modeling Individual Choice. New York: Basil Blackwell, 1989.

Seller, C., J.R. Stoll and J.P. Chavas. "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques." Land Economics, 61(1985): 156-175.

Shonkwiler, J.S. "Double Hurdle Count Data Models for Travel Cost Analysis." W-133 Research Publication, 7th Interim Report, June 1994, pp. 89-103.

Smith, V.K. "Taking Stock of Progress with Travel Cost Recreation Demand Models: Theory and Implementation." Marine Resource Economics, 6(1989): 279-310.

White, H. "Maximum Likelihood Estimation of Misspecified Systems." Econometrica, 50(1982): 1-25.

Table 1. Sample Statistics, N=532.

Lake	Average Number of Visits	Average Number of User Visits	Number Not Visiting
Topaz	1.269	5.921	418
Walker	1.258	5.767	416
Pyramid	.885	6.729	462

Table 2. Single Equation Maximum Likelihood Results.

	Poisson-Log Normal			Negative Binomial		
	Parameter[a]	Coefficient	Std. Error	Parameter	Coefficient	Std. Error
Lake						
Topaz	σ	1.0318	.1096	ϕ	.5555	.2059
	α	2.6752	.5679	α	3.0991	.2963
	β	-.1288	.0277	β	-.1318	.0133
	γ	.0820	.0241	γ	.0805	.0377
	ω	.5878	.0439	ω	.4992	0890
Log likelihood		-533.41			-535.49	
Walker	σ	1.4248	.1345	ϕ	.2060	.1067
	α	1.8914	.1685	α	1.7763	.3812
	β	-.0964	.0087	β	-.0733	.0087
	γ	.1009	.0400	γ	.1069	.0605
	ω	.4241	.0761	ω	.1240	.3112
Log likelihood		-540.72			-553.56	
Pyramid	σ	.9930	.1164	ϕ	.6818	.3114
	α	3.3591	.4724	α	3.4826	.7040
	β	-.0647	.0145	β	-.0608	.0189
	γ	.0191	.0446	γ	.0410	.0658
	ω	.8326	.0226	ω	.8041	.0398
Log likelihood		-399.55			-399.59	

[a] Parameters are: α-intercept
β-own price coefficient
γ-income coefficient (income measured in \$10,000 units)
ω-probability of an impediment to consumption

Table 3. Full Information Maximum Likelihood Results for Three Equation System.

Parameter[a]	Estimated Coefficient	Estimated Std. Error
σ_1	1.3879	.0912
σ_2	1.3018	.0848
σ_3	1.9919	.1040
ρ_{12}	-.2903	.119
ρ_{13}	-.1590	.0637
ρ_{23}	.1218	.0691
α_1	1.8104	.2075
β_{11}	-.1267	.0080
γ	.0787	.0182
β_{22}	-.0921	.0055
β_{33}	-.0551	.0065
ω	.3775	.0337
Log likelihood	-1492.50	

Table 3a. Quasi-Maximum Likelihood Results for Three Equation System.

Parameter	Estimated Coefficient	White's Standard Error
σ_1	1.2493	.0970
σ_2	1.2959	.1302
σ_3	1.9325	.0762
ρ_{12}	-.3837	.1267
ρ_{13}	-.2864	.0814
ρ_{23}	.0151	.0653
α_1	2.3896	.2144
β_{11}	-.1348	.0093
γ	.0741	.0187
β_{22}	-.0947	.0104
β_{33}	-.0554	.0065
ω	.4703	.0296

[a] For subscripts: Topaz = 1, Walker = 2, Pyramid = 3.

Table 4. Welfare Calculations.

Lake	Equivalent Variation	Consumer's Surplus	Compensating Variation
Topaz	-9.620	9.621	9.622
Walker	-12.918	12.921	12.924
Pyramid	-15.056	15.057	15.059

Valuing Travel Time by Isolating the Components of Recreation Benefits

Brian Roach
Division of Environmental Studies
University of California, Davis 95616

Paper presented at W-133 Western Regional Research Conference, Asilomar, CA March 12-15, 1995.

Abstract

The monetary value of travel time is an important unresolved issue in recreation demand analysis. This paper proposes a quasi-experimental design which allows the value of travel time to inferred from data. Theory is developed, in the spirit of random utility models, which does not require the value of time to be linked to an individual's wage rate. The approach rests on the principle that net recreation benefits consist of site quality effects, money costs and time costs. Using a simple two-site two-community design, quality effects are isolated which allows estimation of the value of travel time.

The author would like to thank Doug Larson, John Loomis, and Frank Ward for comments on earlier drafts of this paper.

I. Introduction

While the importance of travel time in recreation demand studies has long been recognized, it remains an unresolved issue (Ward and Loomis, 1986). Along with out-of-pocket travel costs, travel time is a constraint variable in properly specified travel cost and random utility models. High collinearity between travel costs and time often frustrates attempts to specify separate variables. Typically, one assumes a monetary equivalent value of travel time to calculate the cost of visiting a site in terms of a single dollar quantity. However, welfare estimates can be significantly affected by the assigned cost of travel time.

Early work on the value of travel time by Cesario (1976) reviews several commuter studies. He suggests that an appropriate value of travel time is between one-quarter and one-half the hourly wage rate. In a meta-analysis of 77 travel cost studies, Smith and Kaoru (1990) found that travel time was valued at an average of 0.37 of the wage rate. Several theoretical approaches for valuing travel time also attempt to link time value to an individual's wage rate. These models (such as Larson, 1993a; Bockstael, et al., 1987; and McConnell and Strand, 1981) normally assume that individuals tradeoff labor and leisure in a market situation, at least in the long term.

Several attempts have been made to empirically estimate the value of travel time. McConnell and Strand (1981) define a monetarized travel time variable as the after-tax wage rate multiplied by the round-trip travel time in a travel cost analysis of anglers. Their results suggest that travel time is valued at an average of 0.61 times the after-tax wage rate. Larson (1993b) develops a model using monetary travel costs and on-site time to calculate the implied value of time. The average shadow value in the sample was $2.54/hour with some negative values. The empirical analysis by Smith, et al. (1983) considers valuing travel time at both the full wage and one-third the wage rate. They found no clear evidence in favor of either choice.

Several drawbacks are apparent with present approaches for valuing travel time. The proportions suggested by Cesario are based on commuter behavior, not recreational behavior. As suggested by DeSerpa (1971), recreational travel time may be viewed differently than commuting time and enter into utility functions differently. Labor-leisure tradeoff models produce a generic shadow value of time under certain assumptions. Even when the assumptions are valid, one may question whether the average shadow value of time specifically reflects the value of recreational travel time. Several theorists have suggested that links between the wage rate and the value of recreational travel time are difficult to establish (Smith, et al., 1983; Shaw, 1992).

Using wages as a foundation for valuing travel time is problematic for many groups of individuals. The low wages of retired and unemployed recreators will probably be a poor basis for calculating the value of their time (Shaw, 1992). Ward (1983) suggests that the value of time for wealthy people and college students may not be a function of their wage rate. Many people can not marginally exchange time for money in a market situation. McKean, et al. (1995) found that only 19% of visitors to a Colorado reservoir could exchange time for income.

Some economists have suggested using contingent valuation (CV) questions for estimating the cost of travel time. Shaw (1992) recommends supplementing travel cost surveys by asking the willingness to pay for an additional hour of a recreation activity. Wilman (1980) presents a lengthy CV survey which asks a range of questions regarding alternate possibilities and willingness to pay values.

This paper introduces a quasi-experimental design which can infer the value of site- and activity-specific travel time as a function of preferences and socio-economic characteristics. The design attempts to separate the influence of site quality from travel time and costs as factors which affect the utility derived from a recreational visit. Relatively straightforward CV questions are proposed as a mechanism to quantify these differences. Theory is developed which explores the validity of the approach under various assumptions.

II. Theory

This section develops a utility model for an individual visiting a recreation site, similar to the framework for random utility models (Freeman, 1993). The components of utility are categorized into quality effects, money costs, and time costs. The components of utility from visiting one site during a time period are compared to those of another site. Section III proposes a design which allows estimation of each component.

Assume an individual has available time T where T is an individual choice occasion, such as a day. This time may be spent either visiting recreation site 1, site 2, or in alternate leisure activities. Visiting either recreation site will necessitate a fixed travel time and money cost. It is assumed that individuals will not visit both sites during time T. Also, individuals visiting a recreation site may have time left for other leisure activities, such as leisure time at home.

Suppose an individual considers the utility which can be obtained from spending some time at recreation site 1. This utility will be compared to the utility of visiting site 2. Using an additive utility function for illustrative purposes, if some time is spent at site 1 an individual will maximize:

$$U_1 = (\delta_1 V_1^{\alpha_1} - \beta T_1) + (\delta_l L_1^{\alpha_l}) + (\delta_s S_1^{\alpha_s}) \tag{1}$$

Subject to:

$$T = V_1 + T_1 + L_1$$

and

$$M = P_1 + S_1$$

where:

U_1 = Utility for a given time period T where some time is spent at recreation site 1
V_1 = Recreation time spent at site 1
T_1 = Round-trip travel time to site 1
P_1 = Round-trip travel cost to site 1
L_1 = Time spent in alternate leisure activities
S_1 = Monetary savings
M = Stock of money
ß = Value of recreational travel time.

The model assumes that individuals can not trade time for money during time T. Both time and money are fixed and binding. Besides δ_i and α_i, the values of T_1,P_1, and ß are fixed parameters of the model. The value of travel time is assumed constant.

Define a Lagrangian with lambda as the shadow value of money and phi as the shadow value of time. The first-order conditions for maximization are:

$$1.\ \frac{\partial\mathcal{L}}{\partial V_1}=\alpha_1\delta_1 V_1^{\alpha_1-1}-\phi=0$$

$$2.\ \frac{\partial\mathcal{L}}{\partial L_1}=\alpha_l\delta_l L_1^{\alpha_l-1}-\phi=0$$

$$3.\ \frac{\partial\mathcal{L}}{\partial S_1}=\alpha_s\delta_s S_1^{\alpha_s-1}-\lambda=0$$

$$4.\ \frac{\partial\mathcal{L}}{\partial\lambda}=M-P_1-S_1=0$$

$$5.\ \frac{\partial\mathcal{L}}{\partial\phi}=T-V_1-T_1-L_1=0$$

One can solve for S using condition 4 and then solve for lambda using condition 3. The remaining step is to solve for V_1, L_1, and phi using conditions 1, 2, and 5. Using conditions 2 and 5, specify condition 1 as:

$$\alpha_1\delta_1 V_1^{\alpha_1-1}=\alpha_l\delta_l(T-T_1-V_1)^{\alpha_l-1} \tag{3}$$

One can then solve for V_1 in terms of known parameters. An individual will stay on-site until the value of leisure time in some other activity becomes marginally more valuable. The results suggest that the amount of time spent on-site is not a function of the out-of-pocket costs of a visit. Still, increases in the price will reduce utility by decreasing S and eventually cause an individual to not visit the site during time T.

Define the indirect utility of an individual observed visiting site 1 as:

$$U_1^* = \delta_1 (V_1^*)^{\alpha_1} - \beta T_1 + \delta_l (L_1^*)^{\alpha_l} + \delta_s (M-P_1)^{\alpha_s} \qquad (4)$$

This equation is viewed as a generalization of indirect utility from a random utility model (Freeman, 1993). Random utility models consider an individual's utility from visiting a site. The above equation fixes the time period, allowing for time to be spent in other leisure activities.

Next, consider the utility one would get if some time were spent at site 2. A similar utility maximization problem can be defined to solve for V_2^*, L_2^*, and S_2^*. Define the indirect utility for a visit to site 2 as:

$$U_2^* = \delta_2 (V_2^*)^{\alpha_2} - \beta T_2 + \delta_l (L_2^*)^{\alpha_l} + \delta_s (M-P_2)^{\alpha_s} \qquad (5)$$

The difference in indirect utility between visiting site 1 and site 2 can be defined as:

$$\delta_1 (V_1^*)^{\alpha_1} - \delta_2 (V_2^*)^{\alpha_2} + \beta (T_2 - T_1) + \delta_l [(L_1^*)^{\alpha_l} - (L_2^*)^{\alpha_l}] + \delta_s [(M-P_1)^{\alpha_s} - (M-P_2)^{\alpha_s}] \qquad (6)$$

Under certain conditions, this expression can be simplified. First, if the marginal utility of income is relatively constant between $(M-P_1)$ and $(M-P_2)$, then utility can be written as a linear function of the price difference between the two sites. Since the price of a recreational visit is likely to be small compared to M (available stock of money), the assumption of constant marginal utility of money within the price range of the utility maximization problem appears reasonable.

The other simplifying assumption is that leisure time in other activities will be the same for both choice options. This would simplify the utility difference to:

$$[\delta_1(V_1^*)^{\alpha_1}-\delta_2(V_2^*)^{\alpha_2}]+\beta(T_2-T_1)+\lambda(P_2-P_1) \qquad (7)$$

where lambda is the marginal utility of money. Equation (7) specifies the utility difference between visiting sites as three components. The first term represents the perceived quality difference between spending time at the sites. This value will vary across individuals. The second term is the value of the travel time differential and the last term is the price differential.

Under certain conditions, equation (7) would be a reasonable representation of the utility difference. The value of T may be chosen to limit the potential for L. For example, if T is an afternoon and the recreational experience is a boating trip, then the visit and necessary travel should result in a negligible L. If travel time and activities are similar for the two sites, then on may expect that leisure time in other activities will be similar. If T_1 and T_2 are significantly different, it may still be that $L_1=L_2$ because $(V_1+T_1)=(V_2+T_2)$. This may be true if an individual wishes to spend a fixed amount of time away from home.

The utility difference between visiting the two sites can be monetarized through the price differential. If the individual is observed preferring site 1, then P_1 can be increased by some amount, F, to set the utility difference to zero. At a price of (P_1+F), the individual would be indifferent between visiting the two sites. Since utility can be defined in arbitrary units, set the marginal utility of income to 1 so ß is the marginal utility of travel time in dollar terms. At the point of indifference, with $L_1 = L_2$, one can write:

$$[\delta_1 (V_1^*)^{\alpha_1} - \delta_2 (V_2^*)^{\alpha_2}] + \beta (T_2 - T_1) + [P_2 - (P_1 + F)] = 0 \tag{8}$$

or

$$\beta = \frac{[\delta_1 (V_1^*)^{\alpha_1} - \delta_2 (V_2^*)^{\alpha_2}] + [P_2 - (P_1 + F)]}{(T_2 - T_1)} \tag{9}$$

The monetary value of travel time is now stated as a function of the individual components of utility. The next section describes how ß can be estimated using a quasi-experimental design to estimate the quality effect through CV questions while controlling for travel time differences.

III. Quasi-Experimental Design

This section describes a simple scenario for implying the value of travel time. First, consider a pair of substitute sites as described above. A particular choice occasion is considered for a given recreational purpose, such as a fishing day trip to a reservoir. Define two sites available to individuals in community A. Assume that site 2 is further from community A than site 1. It is also assumed that quality for the given recreation experience is higher at site 2 than at site 1. Individuals from community A determine whether the added benefits of higher quality at site 2 offset the additional travel time and costs. As a result of the tradeoff, some individuals obtain more utility from visiting site 1 while others prefer site

2.

Suppose an individual prefers to visit site 1 for a particular recreation experience. The utility difference between visiting sites can be monetarized by the value of F in equation (8). It is proposed that the value of F can be determined using contingent valuation methods. First, a particular recreational activity is described to individuals in community A. Next, individuals preferences for site 1 or 2 are elicited. Then, the value of F asked using a question such as, "If the entrance fee at site 1 were raised X dollars, would you still prefer to visit site 1 or would you switch to site 2?". This type of question does not require an individual to quantify total willingness to pay. Instead, one only need state a willingness to pay differential, which should be easier to conceptualize.

Once the value of F is obtained, most of the terms on the right-hand side of equation (9) can be calculated. Round-trip travel distances and time can be estimated using various computer programs, such as PCMiler software (ALK Associates, 1992). The values of P_2 and P_1 can then be calculated using vehicle operations costs. The unquantified term is the perceived quality difference between time spent at both sites. One way to value this function is to ask additional CV questions. However, such questions would likely be unrealistic and difficult to answer.

A method for isolating the value of the perceived quality difference is proposed by using data from at least one other community. The crucial criteria for the other community is that travel time to both sites is identical. Thus, visitors from the other community will make recreational choices between sites *independent of travel time considerations*. Ward (1983) notes that an unbiased estimate of the value of travel time can be obtained by statistically holding travel time constant. Equal travel time between two sites surely exists in reality for many communities. This allows a quasi-experimental control rather than a statistical control.

Consider a community B for which $T_1=T_2$ for all individuals. The assumption is made that factors which may affect the utility of travel between community B and the sites are constant. For example, the scenery along both trips is assumed to be similar and does enter into the choice process.

The next step is to ask individuals in community B the same CV questions asked in community A. First, their preference for either site 1 or site 2 for a particular recreational activity (such as a day fishing trip) is established. Then, they are asked one or more CV questions to determine their perceptions of the quality difference between the sites.

Suppose an individual in community B prefers site 2. The value of F_b is the minimum fee that would have to be charged at site 2 in order for that individual to switch to site 1. For such a visitor, a utility difference function can be defined similar to equation (8). Since $T_1=T_2$, F_b can be defined as:

$$F_b=(P_1^b-P_2^b)-[\delta_1(V_1^*)^{\alpha_1}-\delta_2(V_2^*)^{\alpha_2}] \qquad (10)$$

Note that the parameters in equation (10) will vary across individuals. If the monetary travel costs are the same, then F_b is interpreted as the monetarized perceived quality difference between the sites. Even if travel costs differ, the monetarized quality difference can still isolated from equation (10).

With a sample from community B, one can statistically estimate the value of F_{bi} (for individual i) as a function of socio-economic characteristics and preferences:

$$F_{bi} = f_i(S_i, M_i, A_i) \qquad (11)$$

A vector of socio-economic characteristics, S_i, could include number of children, age, and education. Income, M_i, may also be an important factor. Attitude scales can be used to elicit a vector of preferences, A_i, for such factors as site crowding, scenic quality, and visitor facilities. An estimated regression for (11) using data from community B allows one to estimate the monetarized quality difference between the sites as a function of measurable variables but independent of travel costs and time.

Next, the proposed approach assumes that the transfer of this regression equation to community A is valid. If the survey instrument in community A collected the values of all necessary independent variables, then the value of the perceived quality difference can be estimated by applying the regression equation (11) in community A. The value of ß from equation (9) can now be estimated for each survey respondent in community A.

IV. Discussion

The method proposed in this article for valuing travel time has several desirable qualities. First, no assumptions are needed about the relationship between the wage rate and the value of travel time. The impact of the wage rate on ß can be empirically tested. If wages influence willingness to pay differentials, then wages will indirectly affect the value of ß. One may find that other socio-economic characteristics, such as the presence of children, have more significant impacts of the value of travel time.

Another feature of the approach is that the value of travel time can be calculated for different activities. For example, visitors interested in fishing may consider travel time less costly than those looking for a place to picnic. Values can also be broken down by length of visit.

The hypothetical nature of the necessary CV questions mentioned in Section III may cause validity problems with the approach. However, the CV questions are relatively straightforward and market-based. Respondents do not need to think about hypothetical quality changes which may be difficult to conceptualize. Instead, they only need to contemplate changes in entrance fees. Well-designed CV questionnaires should minimized potential validity and reliability problems. The theory in Section II suggests that the questions need to be explicit about the description of the recreation experience and available time. Site preferences need to be established for trips of similar duration so the effect of other leisure activities (L) can be assumed to be minimal.

The approach rests on the validity of a transferred equation. A similar issue, the transferability of benefits and demand functions, has been much discussed in the recreation literature (such as Loomis, 1992; and Brookshire and Neill, 1992). The validity of transferred equations can be improved if similar communities are chosen. Average values for various socio-economic variables from published data can be compared to assist in choosing communities.

Communities with equal travel times between two sites may not be common. One possibility is to consider communities with equal travel time between recreation areas of one site. One should also try to control for extraneous influences on the recreational choice process, such as variations in the scenic quality of the trips. The geographical bounds of this design should be relatively homogenous.

Estimating the monetary value of travel time remains an important issue in recreation demand analysis. The method proposed in this paper represents one possible avenue of research on the value of travel time. Conceptualizing different methods which explicitly estimate the value of travel time, such as the approach in this paper, seems to hold promise.

References

ALK Associates. PCMiler (Computer Program). Princeton, NJ. 1992

Bockstael, Nancy E., Ivar E. Strand, and W. Michael Hanemann. "Time and the Recreation Demand Model". American Journal of Agricultural Economics, vol. 69(2):293-302, May 1987.

Brookshire, David S. and Helen R. Neill. "Benefit Transfers: Conceptual and Empirical Issues". Water Resources Research, vol. 28(3):651-655, March 1992.

Cesario, Frank J. "Value of Time in Recreation Benefit Studies". Land Economics, vol. 55(1):32-41, February 1976.

DeSerpa, A.C. "A Theory of the Economics of Time". The Economic Journal, vol. 81(No. 324):828-846, December 1971.

Freeman, A. Myrick, III. The Measurement of Environmental and Resource Values: Theory and Methods. Resources for the Future, Washington, D.C. 1993.

Larson, Douglas M. "Separability and the Shadow Value of Leisure Time". American Journal of Agricultural Economics, vol. 75(3):572-577, August 1993a.

Larson, Douglas M. "Joint Recreation Choices and the Implied Values of Time". Land Economics, vol. 69(3):270-286, August 1993.

Loomis, John B. "The Evolution of a More Rigorous Approach to Benefit Transfer: Benefit Function Transfer". Water Resources Research, vol. 28(3):701-705, March 1992.

McConnell, Kenneth E. and Ivar Strand. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sport Fishing". American Journal of Agricultural Economics, vol. 63(1):153-186, February 1981.

McKean, John R., Donn M. Johnson, and Richard G. Walsh. "Valuing Time in Travel Cost Demand Analysis: An Empirical Investigation". Land Economics, vol. 71(1):96-105, February 1995.

Shaw, W. Douglass. "Searching for the Opportunity Cost of an Individual's Time". Land Economics, vol. 68(1):107-115, February 1992.

Smith, V. Kerry, William H. Desvouges, and Matthew P. McGivney. "The Opportunity Cost of Travel Time in Recreation Demand Models". Land Economics, vol. 59(3):259-277, August 1983.

References (Continued)

Smith, V. Kerry and Yoshiaki Kaoru. "Signals or Noise? Explaining the Variation in Recreation Benefit Estimates". American Journal of Agricultural Economics, vol. 72(2):419-433, May 1990.

Ward, Frank. "Measuring the Cost of Time in Recreation Demand Analysis: Comment". American Journal of Agricultural Economics, vol. 65(1):301-305, February 1983.

Ward, Frank A. and John B. Loomis. "The Travel Cost Demand Model as an Environmenta l Policy Assessment Tool: A Review of the Literature". Western Journal of Agricultural Economics, vol. 11(2):164-178, December 1986.

USING ACTUAL AND CONTINGENT BEHAVIOR DATA WITH DIFFERING LEVELS OF TIME AGGREGATION TO MODEL RECREATION DEMAND

by

Trudy A. Cameron[1]
Department of Economics
University of California - Los Angeles

W. Douglass Shaw
Department of Agricultural Economics
University of Nevada - Reno

and

Shannon R. Ragland, J. Mac Callaway and Sally Keefe
RCG/Hagler Bailly Inc.

[1] A draft of this paper was presented at the 1995 W-133 conference, in Monterey, CA, and for a seminar at the Department of Economics, Florida State University. The authors wish to thank several conference and seminar participants for their comments, and members of the Recreation Work Group from the federal agencies that manage the reservoirs, particularly Matt Rae, Chuck Korsen, Scott Jackson, and John Platt, who provided invaluable assistance in making this research possible. We also wish to thank Jeff Englin, John Loomis, George Parsons, Bob Berrens, and Frank Ward for their help on the survey design and modeling issues. Preparation of this paper supported by the Nevada Agricultural Experiment Station.

USING ACTUAL AND CONTINGENT BEHAVIOR DATA WITH DIFFERING LEVELS OF TIME AGGREGATION TO MODEL RECREATION DEMAND

1.0 INTRODUCTION

Public concern about anadromous fish species in the Columbia river system led in recent years to consideration of changing water levels behind the dams to enhance the ability of fish to migrate. Even without such consideration, changes in reservoir operations are possible under impending renewal of regional and international hydropower agreements. In both cases, federal operating agencies must comply with National Environmental Policy Act (NEPA) procedures, including preparing an Environmental Impact Statement (EIS) and are therefore studying the effects of water level changes on water-based recreation at several federal reservoirs and controlled downstream river stretches that are located throughout the Columbia River Basin. Our fundamental objectives are to:

- predict how often individuals take trips to **each** of several federal waters (referred to as projects) under existing and hypothetical water levels
- predict recreation values and subsequently predict changes in those values for changes in water levels at each of the projects.

We derive these values in the context of a modern recreation demand, or travel cost model (TCM).[2] Some studies investigate reservoir management issues that are similar to the ones faced here (Cordell and Bergstrom; Florida Dept. of Environmental Protection; Ward et al. 1989). In our approach we incorporate the project's water level as a key destination characteristic (Morey), allow for substitution (Kling), use contingent behavior (Cameron 1991), and correct for possible bias from individuals who do not respond to the survey or do not take any trips to recreation destinations under consideration. Our model is made more

[2] For recent reviews of recreation demand modeling see Bockstael, McConnell and Strand (1991).

appropriate for application to the general recreator population by combining census data for origin 5-digit zip codes with our sample data.

We combine data that has a time dimension to it with data that vary across individuals and projects (known as cross sectional data). We use data on actual recreation trips for May, June, July, and August of 1993 combined with actual water levels for each of these months. Few recreation studies we know of attempt to combine this "time series" data with cross sectional data in a model (an exception is Cole et al.).

2.0 SAMPLE, DATA, AND KEY MODELLING CONSIDERATIONS

The recreation model we develop is used to estimate monthly demand for nine specific federal waters in the Columbia River basin. We estimate the model using data collected using a mail survey questionnaire mailed in the fall of 1993.[3] The sample includes recreators, as well as members of the general population of the Pacific northwest (PNW), whose addresses were obtained from a telephone directory. It also includes an oversampled pool of residents from counties adjacent to the federal waters considered for the analysis, recreators intercepted while at the federal waters and asked to participate in the study by mailing in a postcard containing their address, and willing volunteers from an earlier survey effort (Callaway et al. 1993). Defining the sample this way may lead to possible biases, and these are considered in Section 4.1, below.

The data collected include information on the household's 1993 actual water-based recreation trips and the regional destinations for those trips, specific information about the household's typical trip to regional projects, responses to two hypothetical behavior

[3] Details are provided in a report (Callaway et al 1995) or in a much longer version of this paper (Cameron et al. 1995).

questions, and the usual household information.

As compared to other water-based recreation, there are several unusual features of Columbia River basin recreation at these federal waters. These, and our handling of these features, are as follows.

Water Level Data Problems

Some of the water levels that will be analyzed represent drastic departures from the water levels that prevailed during 1993 (our survey season). Observed water levels, measured as monthly average summer levels at various projects, are used to explain monthly demands. However, these water levels often move up and down together. Collinearity between water levels at some projects is so strong that using the actual trip data alone resulted in very serious multicollinearity problems.[4] We resolve the difficulties associated with the observed water level data by including *contingent behavior* (CB) questions (Cameron 1991; Englin and Cameron 1994) in the survey, coupled with computer-enhanced photographs and graphical and verbal depictions of possible water level changes.[5] An individual is allowed to state that he would, would not, or does not know if he would take a different number of trips to each regional project under the hypothetical, *versus* the actual water levels in 1993. If he would take a different number of trips, he is asked how many more or fewer trips he will take to each regional water. This increases the amount of independent variation in the water level data because no two projects were treated identically on two or more survey versions.

---

[4] As a preview, this multicollinearity was evidenced by its classic symptoms, namely drastically changing parameters as alternative water levels are dropped in and out of the model, and insignificance in the "own" water level when accompanied by these alternatives. Further, simple correlation coefficients showed evidence of a strong linear positive or negative relationship between water levels in several instances. For example, the correlation coefficient for actual water levels at Albeni Falls and Hungry Horse is .98. Using the hypothetical levels posed in the questions for version 1 and coupled with the actual water levels decreases this correlation coefficient to .26.

[5] A copy of the survey insert, including the color computer-enhanced photographs is available on request. Thanks go to Matt Rae and other key *ACE* members for these photographs.

Large Number and Variety of Recreation Destinations

Our model must be able to predict the total number of trips to <u>each</u> destination (whether reservoir or river), so that reallocations of visits due to water level changes can be examined at each federal project. All trips for boating, shore use, etc., not just trips for a given recreation activity like fishing, must be considered.

We handle the substitution between projects by assuming that survey respondents choose a destination from a set of nearby projects, with the water level at the destination and other projects taken into consideration according to their importance in choosing a destination. We collect trip information to each of the nearby projects and to "all other" waters in the same region as the individual's residence.[6] The responses are pooled together across regional versions of the survey so that the demand for a project can be estimated as a function of responses and characteristics of all the individuals in the sample who had an opportunity to report a trip to that project. Last, to accommodate different types of recreation, we use intercept dummy variables for the type of recreator each individual appears to be (holder of fishing license, boat owner, or both).

Sample versus General Recreator Population

Our sample contains individuals who were randomly selected from the general population because we wanted to address the issue of "who" was in the set of impacted individuals right from the start. Some individuals did not return their surveys, and the nonrecreators in our sample may have been less likely to return the survey than people who visited the projects. We adjust for differences between the general population and those individuals who return the survey by combining 5-digit zip code census data for origins with

---

[6] An earlier survey effort indicated that the vast majority of trips that are taken by those that live in the Columbia River basin are to nearby, or regional waters.

the rest of the data to estimate the probability that an individual responds to the survey. This probability is then used to make a rudimentary selectivity correction in the recreation demand models via incorporating the usual Inverse Mills Ratio (*IMR*) (Heckman 1976).

3.0 THE THEORETICAL DEMAND MODEL

The approach taken here is a two stage heteroscedastic model of average summer season (May through August) monthly demand (trips) for a project, corrected for survey response bias.[7] A separate demand equation is estimated for the remainder of the season. This rest-of-year demand is incorporated solely in order to "complete" the set of disaggregated monthly demands and thereby to facilitate combining the monthly data and annual data employed in this study. It is not considered to be of independent policy interest.

3.1 The Basic Model

The model is laid out with the following notation:

Let X = vector of individual-specific socioeconomic determinants of demand (including travel costs) that *do not vary over time* during the summer months (see Appendix, available from the authors, for details on travel costs or prices)

Z_t = vector of socioeconomic or other determinants of demand that *do vary over time* during the summer months

W_t = vector of monthly water levels at all nine main waters in each of May - August.

Let monthly demands (q^t) be expressed as:

$$q_t = X'\beta_x + Z_t'\beta_z + W_t'\beta_w + \epsilon_t, \quad t = \text{May, June, July, August} \tag{1}$$

[7] A similar approach to ours, with a focus on water level changes at several different reservoirs was implemented by Ward (1989). In Ward's application however, the water level change modeled is a *total removal* of water, which is actually simulated by changing the site price until zero visits occur at the site. Our approach differs mainly because we examine less severe reductions (and increases) in water levels using a water level variable within the model.

Let rest-of-year demand (q_R) be expressed as:

$$q_r = X_r'\gamma_x + \epsilon_R \tag{2}$$

This specification is used because off-season water levels are not available. Actual annual 1993 demand (q_A) can be expressed as:

$$\begin{aligned}
q_T = \ & (X'\beta_x + Z_5'\beta_z + W_5'\beta_w) + && \text{monthly demands} \\
& (X'\beta_x + Z_6'\beta_z + W_6'\beta_w) + && \\
& (X'\beta_x + Z_7'\beta_z + W_7'\beta_w) + && \\
& (X'\beta_x + Z_8'\beta_z + W_8'\beta_w) + && \\
& (X_r'\gamma_x) \quad + && \text{off-season demand} \\
& \epsilon_T, \ \ T = A, B, C && \text{annual residual}
\end{aligned} \tag{3}$$

For T = A, we have actual annual demand. We also have analogous contingent annual demands from the two contingent behavior questions, denoted q_B and q_C.

3.2 Adjustments in the Basic Theoretical Model

Note that we do not have trip or water level information on a month by month basis for any but the summer months. To be able to express annual demand as a function of *average summer water levels* we divide the summer water levels by 4, or:

$$(W_5 + W_6 + W_7 + W_8)/4 \tag{4}$$

Because of this data availability, we need to use $q_A/4$ for actual 1993 annual demand (and $q_B/4$ and $q_C/4$, analogously).

Annual demands can therefore be expressed as a function of $\Sigma W_t/4$

$$q_A/4 = (\Sigma X/4)'\beta_x + (\Sigma Z_t/4)'\beta_z + (\Sigma W_t/4)'\beta_w + (X_r/4)'\gamma_x + \epsilon_A/4 \tag{5}$$

or

$$q_A/4 = X'\beta_x + (\Sigma Z_t/4)'\beta_z + (\Sigma W_t/4)'\beta_w + (X_r/4)'\gamma_x + \epsilon_A/4 \tag{6}$$

The simplification of the first term is possible because X is time-invariant.

The same holds for $q_B/4$ and $q_C/4$. Thus, the same parameters, β_x, β_z, β_w, γ_x can be made to appear in all eight demand equations available for each respondent.[8] Arrayed similarly, the correspondences between the parameters are clear. The four monthly observations are:

$$\begin{array}{lllllll}
q_5 & = & X'\beta_x + & Z_5'\beta_z + & W_5'\beta_w + & 0'\gamma_x + & \epsilon_5, \\
q_6 & = & X'\beta_x + & Z_6'\beta_z + & W_6'\beta_w + & 0'\gamma_x + & \epsilon_6, \\
q_7 & = & X'\beta_x + & Z_7'\beta_z + & W_7'\beta_w + & 0'\gamma_x + & \epsilon_7, \\
q_8 & = & X'\beta_x + & Z_8'\beta_z + & W_8'\beta_w + & 0'\gamma_x + & \epsilon_8.
\end{array}$$

The rest-of-year observations are:

$$\begin{array}{lllllll}
q_R & = & 0'\beta_x + & 0'\beta_z + & 0'\beta_w + & X_r'\gamma_x + & \epsilon_R.
\end{array}$$

and the three annual observations are:

$$\begin{array}{lllllll}
q_A/4 & = & X'\beta_x + & (\Sigma Z_t/4)'\beta_z + & (\Sigma W_t/4)'\beta_w + & (X_r/4)'\gamma_x + & \epsilon_A/4, \\
q_B/4 & = & X'\beta_x + & (\Sigma Z_t/4)'\beta_z + & (\Sigma W_t/4)'\beta_w + & (X_r/4)'\gamma_x + & \epsilon_B/4, \\
q_C/4 & = & X'\beta_x + & (\Sigma Z_t/4)'\beta_z + & (\Sigma W_t/4)'\beta_w + & (X_r/4)'\gamma_x + & \epsilon_C/4.
\end{array}$$

The fact that summer monthly plus rest-of-year demands must sum to annual demand places strong restrictions on viable functional forms for the demand equations: they must be linear in q and linear in parameters. Note also that the information in $q_A/4$ is redundant with that in q_5 through q_8 plus q_R, so q_A data will be dropped from the estimating models.

[8] Some respondents, who declined to answer the contingent behavior questions, have only six pieces of demand information each.

4.0 EMPIRICAL MODEL

In this section we discuss how the theoretical model is recast so that it can be estimated. To begin we discuss corrections for nonresponse bias. Following this, we discuss the empirical specification for the two stage recreation demand model.

4.1 Survey Response/Non-Response Model

This section addresses potential non-response bias. We correct for selection bias by respecifying the recreation demand model to reflect the fact that we only have information for the individuals who returned the survey. We do this by adding two preliminary probit models which explain the probability that the individual returned the survey questionnaire, estimating the *IMR*, and including this variable in the recreation demand model.

This response/non-response model estimates the probability that an individual returns the survey as a function of:

- 5-digit zip code data from the 1990 Census--merged according to the zip codes on the full intended sample.
- distance data from intended sample zip codes to the various waters considered on each version of the survey
- a variable that indicates the survey sample strata from which the individual is drawn.

This survey return model yields the response/nonresponse IMR:

$$\phi(Y_i'\alpha) \; / \; [1-\Phi(Y_i'\alpha)] \tag{9}$$

where Y_i is a vector of the explanatory variables (these are enumerated in Table 1a and 1b), and Φ and ϕ are the cumulative distribution function and the probability density function for the standard normal evaluated at the values of the independent variables.

We use the IMR variables from the response/nonresponse models for the actual and the contingent demand information as additional explanatory variables to control for heterogeneity in the propensity to return the survey. As will be seen, these are sometimes important determinants of recreation demand. We have not before seen a more complete attempt to adjust for survey return bias in a recreation demand model. As such, we have made an attempt to control for the bias that stems from the sample strata the individual is from, as well as other factors that contribute to their returning the survey.

4.2 Two-Stage Empirical Specification

We estimate project demand in two stages. In the first stage the probability that the individual recreator takes positive trips to a particular project j is estimated. Because a recreator can take trips somewhere other than to any one particular project, his zero trips will not enter into the second stage. In the second stage, the continuous model of trip demand for project j is estimated, conditional on an individual having taken a trip to project j. This specification is somewhat like a common maximum likelihood estimation (MLE) Tobit Model, generalized to allow for two different indexes:

- $G'\beta_g$ is the G vector of variables and their parameters explaining the zero-trips/positive-trips choice
- $H'\beta_h$ is the H vector of variables and their parameters explaining the number of positive trips

However, we are not estimating a tobit model in the sense that negative predicted trips would have their densities moved to, or placed at zero. The probability of taking a positive number of trips is $\Phi(G'\beta_g)$. Thus, the IMR for positive trips is:

$$\lambda^G = \phi(G'\beta_g) / [1-\Phi(G'\beta_g)] \tag{10}$$

The expression for expected trips, given that positive trips are taken is $H'\beta_h$. The expression for unconditional expected trips is therefore:

$$\Phi(G'\beta_g)\ [(H'\beta_h + \rho\sigma\lambda^G)] \tag{11}$$

For the $G'\beta_g$ index that determines zero versus positive trips, the variables and parameters in the model are: $G = (X, Z_t, W_t, X_r)$and $\beta_g = (\beta_x, \beta_z, \beta_w, \gamma_x)$. Descriptions of the elements of the variable vectors are presented in Table 2. **X** may include the price (travel cost) to the own project and alternatives, and income. Z_t includes a summer month dummy variable and an index of the tendency to take trips in some particular months, calculated using average trips from data for the northeastern U.S.[9]

For the monthly observations:

$$G'\beta_g = X'\beta_x + Z_t'\beta_z + W_t'\beta_w + \epsilon_t \qquad \epsilon_t \sim N(0,\sigma_t^2) \tag{12}$$

And, for the rest-of-year observations:

$$G'\beta_g = X_r'\gamma_x + \epsilon_R \qquad \epsilon_R \sim N(0,\sigma_R^2) \tag{13}$$

Similarly, for the annual observations:

$$G'\beta_g = (\Sigma X/4)'\beta_x + (\Sigma Z_t/4)'\beta_z + (\Sigma W_t/4)'\beta_w + (X^r/4)'\gamma_x + \epsilon_T) \tag{14}$$
$$\epsilon_T \sim N(0,\sigma_T^2)$$

For the $H'\beta_h$ index that determines number of trips, given trips are positive, the expressions are analogous. This specification allows $H \not\equiv G$, $\beta_h \neq \beta_g$, which is more general than would be true in the traditional tobit specification.

[9] We thank Dr. George Parsons for providing estimates of the total number of trips by month from his New England recreation data set. We use these estimates to proxy the unobservable tendency to take a trip in June, July, etc., and note that this variable will not be correlated with water levels in the Columbia River basin.

Correction for Heteroscedasticity

Note that the error terms will be heteroscedastic due to the presence in the estimating specification of data at three different levels of time-aggregation: monthly, rest-of-year, and annual. The correction for heteroscedasticity is fairly general when viewed in the context of the estimation method, and both are discussed in the next section.

5.0 ESTIMATION METHOD

Though we are not estimating a Tobit model, it will facilitate discussion of the estimation method to review a conventional Tobit log-likelihood function under homoscedasticity. Let $I_i = 1$ if $q_i > 0$, $I_i = 0$ if $q_i = 0$. With a single index this function is:

$$\max_{\beta_g, \sigma_g} \quad \log \mathcal{L} = \sum_{I=0} \log\{1 - \Phi(G'\beta_g/\sigma_g)\}$$
$$\sum_{I=1} -(1/2)\,\{\log(2\pi) + \log \sigma_g^2 + [(q_i - G'\beta_g)^2/\sigma_g^2]\,\} \qquad (15)$$

Heteroscedasticity across the three different data types (i.e. monthly, rest-of-year, and annual) and the use of two different indexes, $G'\beta_g$ and $H'\beta_h$, complicate the estimation so that we cannot use the above equation to estimate the model. Instead, we use a two-stage Heckman-type model.[10] The two-stage estimation process is:

[10] As will be seen below, we are investigating the differences in use of a full-information maximum likelihood approach.

STAGE 1: Heteroscedastic Probit (monthly, rest-of-year, annual variances)

$$\max_{\beta_g,\delta_r,\delta_T} \log \mathscr{L} = \textstyle\sum_i^t I_i \log\{ \Phi(G'\beta_g) \} + (1 - I_i) \log\{ 1 - \Phi(G'\beta_g) \}$$
$$+ \textstyle\sum_i^r I_i \log\{ \Phi(G'\beta_g/\exp(\delta_r)) \} + (1 - I_i) \log\{ 1 - \Phi(G'\beta_g/\exp(\delta_r)) \}$$
$$+ \textstyle\sum_i^T I_i \log\{ \Phi(G'\beta_g/\exp(\delta_T)) \} + (1 - I_i) \log\{ 1 - \Phi(G'\beta_g/\exp(\delta_T)) \} \quad (16)$$

where Σ_i^t signifies the sum over all monthly observations, Σ_i^r signifies the sum over all rest-of-year observations, and Σ_i^T signifies the sum over all annual observations, both actual and contingent. The error standard deviation for the monthly data is normalized to unity (or, equivalently, β_g is actually β_g^*/σ_g). Defining the indicator variables $D_r = 1$ for rest-of-year data, 0 otherwise, and $D_T = 1$ for annual data, 0 otherwise, allows the index to be generalized to $G'\beta_g/\exp(\delta_r D_r+\delta_T D_T)$. From this, we save the fitted inverse Mill's ratio, $\lambda_G = \phi(G'\beta_g/\exp(\delta_r D_r+\delta_T D_T)) / [1-\Phi(G'\beta_g/\exp(\delta_r D_r+\delta_T D_T))]$.

STAGE 2: Use LIMDEP heteroscedastic Tobit algorithm on <u>only</u> those observations with positive trips (equivalent to OLS estimated by maximum likelihood[11]):

$$\max_{\beta_h,\sigma,\delta_r^*,\delta_T^*} \log \mathscr{L} = \textstyle\sum_{I=1} -(1/2) \{\log(2\pi) + \log(\sigma^2\exp(\delta_r^* D_r+\delta_T^* D_T))$$
$$+ [(q_i - H'\beta_h)^2/(\sigma^2\exp(\delta_r^* D_r+\delta_T^* D_T))] \} \quad (17)$$

where H now includes λ_G (interacted with dummies for monthly, rest-of-year, and annual data types), and the usually constant σ^2 is generalized to differ across the three data types to

[11] While there are no I=0 limit observations in the log-likelihood function in (17), we use the tobit procedure in *LIMDEP* for the second stage because this algorithm conveniently allows for heteroscedastic errors and permits us to take advantage of the higher-level language of LIMDEP. The one problem with relying on this packaged algorithm is that the LIMDEP output for this second stage reports t-test statistics that do not correct for the presence of estimated regressors (the λ_G variables). While we report the t-statistics from the LIMDEP output, we note that these are derived from a variance-covariance matrix that is not strictly correct.

accommodate the fundamental heteroscedasticity in our data.

It is straightforward to program a full information maximum likelihood algorithm that allows simultaneous estimation of the two sets of slope and intercept parameters, β_g and β_h, as well as the conditional heteroscedastic error variance parameters, δ_r, δ_T, σ, δ_r^*, and δ_T^*, as well as the correlation between the latent probit dependent variable and the observed continuous trips variable (given trips are positive). This correlation parameter is ρ. We have programmed such a log-likelihood function for the case of Water 7 (Lower Granite Lake), using the converted two-stage point estimates as starting values. However, we have not yet managed to achieve convergence in this optimization.[12]

6.0 DERIVATION OF WTP (APPROXIMATE CONSUMER'S SURPLUS)

An individual's WTP to bring about a change in water levels can be defined in terms of expected consumer's surplus.[13] The formula for consumer's surplus for an individual facing a change in water levels is:

$$E[CS] = \int_{W_0}^{W_1} Q^*(P|\epsilon)dW$$

where Q^* (.) is the observed demand at initial water level W_o and, conditional on ϵ, and W_1 is the water level after the change.

Because of complexities associated with actually calculating this E[CS] for each individual and for every water level change that needs to be considered, we actually

[12] This algorithm ran under GQOPT on a UNIX system. The initial DFP portion of the optimization, with a convergence criterion of 10^{-6} ran for an elapsed time of 8 days without convergence, although these were "good" iterations. We are exploring reduced specifications with the intent of assessing the distortions in the parameter variance-covariance matrix from using the two-stage method. It is unlikely that we will be able to pursue FIML estimates for all parameters for all specifications in this study with our current computing hardware.

[13] It is expected consumer's surplus because of the stochastic nature of ϵ. This is not uncommon in modern recreation demand modelling (eg. Hellerstein 1992).

approximate E[CS] for a given water level by estimating the area under *unconditional expected trip demand function* from individual observed price up to individual choke price.

To derive E[CS] for a change in water levels, we do this over again for a different water level and subtract the difference between the two areas to estimate the E[CS] for the water level change. The choke price is determined by solving the portion of the demand formula not including the inverse Mill's ratio terms for the price axis intercept (the choke price will be at least this much) and then switching to the full demand formula and incrementing price upwards in intervals of .001 until quantity demanded becomes negative.[14]

Recall from above equation that unconditional expected trips are:

$$E[q] = \Phi(G'\beta_g/\exp(\delta_r D_r+\delta_T D_T))\ [(H'\beta_h + \rho\sigma_t\lambda_t^G + \rho\sigma_r\lambda_r^G + \rho\sigma_T\lambda_T^G)\ ] \quad (19)$$

The own-price of a trip (travel cost) appears non-linearly in several places in the expression for E[q], namely in the vectors G, H, and the IMR, and in each of the λ^G terms. There are three different $\rho\sigma$ coefficients on the three λ_G terms, because the value of σ, although not identified by the second stage parameter values, will differ across the three heteroscedastic data types.

Hellerstein (1992) notes that numeric techniques are required to solve the integral required for E[CS]. This is due to the fact that Φ and ϕ have complex mathematically nonlinear forms that depend on the values of the parameters and the independent variables.

---

[14] It is technically possible to solve for the "zeros" of the demand function for each individual, using an algorithm that minimizes the square of quantity demanded as a function of price. However, this strategy proved too computer-intensive to allow simulations to be completed in any reasonable amount of time. We have settled for a close approximation to the true choke price for each individual (based on their estimated demand function). Since all individual demand functions are constrained to have the same slope with respect to price, this approximation should be comparably close for all observations.

An explanation of how we dealt with this issue is elsewhere (Cameron et al. 1995 or Callaway et al. 1995).

7.0 RESULTS I: COEFFICIENT ESTIMATES

We obtained coefficients for the basic probit model of survey response, the probit models for the participation decision at each of the nine projects, and the continuous heteroscedastic models for each of these nine projects. Because presentation of all of these coefficients would be tedious (there are four sets of probit survey response results and pairs of model estimates for each of the nine projects), we summarize the results below.[15]

7.1 Survey Response/Non-response Model

Tables 1a and 1b give the parameter estimates for the two probit models that were estimated for each of the four regional survey versions. The models reported in Table 1a capture the effects of different sociodemographic, distance, and Census zipcode data upon the respondent's probability of responding with revealed preference information sufficient to allow their responses to be included in the estimating sample. Table 1b uses identical variables to explain the probability of responding with contingent preference information sufficient to allow these responses to be included in the estimating sample. Two separate probit models were estimated for each region because noticeably more respondents provided revealed preference than contingent preference data. We assume, for expedience, that the processes leading to actual (versus contingent) response completing are independent. (In reality, they are probably not.)

Scrutiny of Tables 1a and 1b will reveal that we have not strived to determine the

[15] All coefficient estimates are available in Callaway et al. 1995.

most parsimonious specification for each probit model. Multicollinearities do exist between some of the explanatory variables, so we are not relying entirely upon individually statistically significant t-ratios to drive our specifications. Maximizing "fit" is relatively more important in this context. Variables that tend to be significant determinants of survey response included the distance to the project, the survey sample strata group, and various demographic variables. In cases where we had priors on the sign of the variable, all variables had the expected influence on the probability of returning the survey questionnaire, except in a few instances. For some of the distance variables, the further the respondent resides from that particular water, the more likely the individual was to return the completed survey questionnaire.[16] One possible reason for this counter-intuitive result is that being far from one thing usually means being close to something else. Being far from some of our waters is correlated with being close to some heavily populated areas (most likely the Portland, Oregon area or Seattle, Washington area). Distance from some waters could be correlated with other characteristics that increase or decrease the probability of survey response in a manner not captured by our other control variables.

Several variables associated with income and affluence were examined in these models. In the Region 4 survey response model, the proportion on public assistance income had a negative influence on returns, but its influence was not individually significant in the other models. Median zipcode income was significant in three of the four survey response models, but had a negative influence on return in two of these models and a positive influence in one. These findings may simply be an artifact of multicollinearity among the various sociodemographic variables. It is plausible that high income individuals have too

[16] Note that the identities of waters 1 through 4 (if applicable) differ across regions in Tables 4a and 4b. We could assign each water a number from 1 to 9 as in most of this study, but this would increase the empty space in these two tables.

high a marginal value of time to return the survey, but also possible that high income individuals are more likely to take interest in an outdoor recreation survey. Last, the median house value, which may be correlated with real wealth, had a negative influence on returns in versions 2 and 3.

Of the urban composition, racial and ethnicity, education, and other socio-demographic variables tried, various ones had some significance in some of the models, but there is no overwhelming clear trend in these.

The *IMRs* derived from these probit models (appropriate to each type of data--actual or contingent) are included as additional explanatory variables in both stages of the recreation demand model below. These "estimated regressors" allow estimation of a coefficient that is the product of (i.) the error correlation between the sample selection process and the demand equations conditional on presence of an observation in the estimating sample and (ii.) the error standard deviation in the estimated demand equations. We are not able, with current estimation technology, to readily estimate these preliminary sample selection probit models simultaneously with the two parts of the demand model. To do so would require algorithms to evaluate a multivariate normal cumulative density. While simulated moments techniques are available, our specifications are rather too highly parameterized to make such estimation economical.

7.2 The Demand Model: Heteroscedastic Probit First Stage

Our demand modelling utilized a sample limited only to respondents who reported taking at least some trip to some water during the recent season. This could be one of the Federal projects or any one of the "other waters." Thus the modelling exercise is limited to current active users of surface waters in the Pacific Northwest. Given that almost no respondent who did not take trips during the season under actual conditions was induced to take trips under the contingent scenarios, it was judged unlikely that modelling the overall participation/nonparticipation decision would be very informative. Instead, for this group of general water users, we focus on the distribution of their trips among the different waters in the regional choice set. Any specific federal project will have many water recreators with zero trips to that particular water, but these respondents will have reported at least one actual or contingent trip to some other water.

Tables 3a, 3b and 3c present the results of the demand modelling exercises. Each sub-table gives estimates for each of three waters (Table 3a: Waters 1,2,3; Table 3b: Waters 4,5,6; and Table 3c: Waters 7,8,9). For each water, there are two columns of results. The first column is a probit model to predict whether the respondent took any trips to that water. The second column is the Tobit portion of the model, to be discussed below. The basic variables that were considered as explanatory variables are listed and defined in Table 2. The key results here are that the own prices were negative and significantly different from zero in all models except a few, and the own water level variable was most often positive and significantly different from zero. The cross price and cross water level terms are mixed in sign and significance, painting no clear picture of the importance of alternatives in each of the project participation models that is estimated. For example, for the Lake Roosevelt participation model, Pend Oreille and Lower Granite may be substitutes, as indicated by the cross price terms, but if the water level at Pend Oreille increases, participation at Grand Coulee also increases. This unusual effect may be due to collinearity between actual water levels at each water. The actual historical water levels are sometimes highly correlated across waters. The contingent scenarios untie these correlations in some instances but not in all. To have used the contingent scenarios to completely orthogonalize the various water levels would have been extremely helpful to the empirical analysis, but the costs of the computer simulated color photographs and the logistics of multiple survey instruments for each region hindered our ability to achieve this goal.

The non-price and non-water-level variables--income, the water-based activity dummy variables (fishing license, boat ownership, or both), the seasonal visitation rate control variable--are most often of the expected sign, but the significance of them varies from water to water.

7.3 Demand Model: Heteroscedastic Tobit Second Stage

The set of candidate explanatory variables in the second stage models are essentially the same as for the first stage models, except we include the λ_t, λ_r, and λ_T inverse Mill's ratio terms from the first stage heteroscedastic probit participation model. These results are again quite mixed. Of most interest is that while the prices and own water levels often have the expected sign and are significantly different from zero in the first stage models, these variables are seldom significantly different from zero in the second stage models. Own-price is negative and significant in the Pend Oreille and Hungry Horse demand equations only. This indicates that the major influence of these variables may be in the participation decision itself; once an individual decides to visit a water, he pays little attention to the price and water level in determining the frequency of his monthly visits.

The cross price and cross water level terms are also again mixed in sign and significance. The usual expectation is that of substitution among waters, and the models indeed identify some substitutes. A negative price, or positive water level coefficient on an alternative water may indicate some complementarity. Such complementarity is unlikely unless it is an artifact of multiple-site trip taking, but we do not distinguish between single and multiple-destination trips in our models.

8.0 RESULTS II: EXPECTED TRIPS AND CONSUMER'S SURPLUS

Expected average trips and consumer's surplus can be calculated for baseline actual 1993 water levels and then for any change from these baseline levels to some hypothetical water level. They can also be estimated for any set of water levels that one wishes to examine. As part of the overall project for the federal agencies, we considered several of these hypothetical levels, usually pegged to some system operating conditions or "strategies" (SOSs) that might be part of the scheme to flush the salmon smolts out to the ocean. We examine two strategies below, a "recreation" and "fishery" strategy. The former essentially tries to enhance recreation opportunities by filling reservoirs by the end of June, maintaining the pool at full through the end of August. These conditions are more or less considered to be "optimum" recreating conditions at reservoirs. The latter strategy is aimed at assisting downstream fish migration and enhancing flows for spawning.

8.1 Expected Trips and Changes in Expected Trips

For each of the nine projects, we estimate expected monthly trips for a given water level. For the nine projects, the predicted expected average monthly number of trips under conditions in 1993 (our "baseline") is sometimes quite small, for example varying from .5 of a trip in May to 1.06 trips in August for Pend Oreille. For all the projects, expected average trips are lowest at John Day (.09 in July) and highest at the Kootenai River (1.7 in July).

Comparing expected trips under the recreation and fishery strategies, one basically can see the average expected trips decline at all nine projects, as would be expected. For example, assuming water levels are consistent with the average levels over the past 50 years, but controlled to enhance recreation, the expected average trips at Hungry Horse in June is .807. Under the same conditions except changed to the fishery strategy, expected trips fall to .44, about half.

The number of expected trips falls with changes in water levels under the fisheries strategy, but in some cases not by a significant amount. As another example, expected trips at Dworshak Lake for the recreation strategy, again otherwise assuming the 50-year average water level conditions, is 1.367 for July, and 1.186 under the fisheries strategy.

8.2 Expected Consumer's Surplus

Average expected consumer's surplus is calculated first for the baseline actual water level in 1993. The average expected monthly consumer's surplus for baseline conditions in 1993 can be interpreted as the expected monthly WTP rather than do without the water, given the 1993 water levels. For all nine projects, this number is reasonable, varying from about $13 (each summer month) for Lake Koocanusa, to $99 (August) for Lake Roosevelt. These monthly welfare amounts cannot easily be compared to other welfare calculations because those are usually annual or "per-trip" measures. A range of the value "per outing" for water-based recreation is from about $20 to $60, with some estimates being higher for recreation such as fishing for salmon in Alaska. This would suggest that our estimates of monthly WTP are low because of course, many expected trips are predicted per month. However, many of the per-trip or per-day values are not estimated in a probabilistic model, and thus are not "expected" welfare measures; we often know in other models that the individual actually took a trip.

We also estimated average consumer's surplus for the recreation and fisheries strategies in order to contrast these. Assuming the 50-year average levels pertain, the July average expected consumer's surplus for Hungry Horse under the recreation strategy is approximately $72. Under the fisheries strategy, this falls to $40, slightly more than half the monthly WTP under optimum recreation conditions. For other projects at other times during the summer, this change is not so dramatic, which is due to different estimated demands, as well as different water level conditions for specific projects.

9.0 CONCLUSIONS

This research demonstrates that making a rigorous attempt at correcting for survey response bias seems to matter a good deal in models for most of the destinations we considered. In addition, we show that there may be an advantage in contingent behavior responses over contingent valuation ones in sorting out answers that are indicative of problems on the part of the respondent (see Cameron et al. 1995). In the context of modelling river basin changes, contingent behavior responses may be critical information, as there is so much multicollinearity between water levels at various places in the basin.

We have offered a model that explains how behavior and values change in response to a wide variety of water level changes. At the outset, it is not absolutely clear whether water levels at reservoirs really matter in determining participation at, and frequency of trips taken to various federal reservoirs and rivers in the Columbia River basin. Based on our analysis of the data collected for this study, we conclude that water levels at a water (the "own water level") do strongly contribute to the probability that an individual will visit a federal water at all. Perhaps of not great surprise, this influence diminishes in the model that explains the frequency of trips taken.

REFERENCES

Bockstael, N.A., T.E. McConnell, and I. Strand. 1991. "Recreation." In J.B. Braden and C.D. Kolstad, eds. *Measuring the Demand for Environmental Quality.* Amsterdam: Elsevier Science Publishers B.V. pp. 227-270.

Callaway, J.M. et al. 1995. *Columbia River Systems Operation Review Recreation Impacts: Demand Model and Simulation Results.* Final report prepared for Mathew Rae and the Recreation Work Group, U.S. Army Corps of Engineers, Portland, Oregon. RCG/Hagler Bailly, Inc. (May)

Callaway, J.M., W.D. Shaw, and S. Ragland. 1993. *Estimation of Recreation Impacts for the Columbia River System Operation Review: Methods for Phase II.* Final report prepared for the Recreation Work Group of SOR. RCG/Hagler Bailly, Inc.

Cameron, T.A., W.D. Shaw, J.Callaway, and S. Ragland. 1995. Recreation demand modelling using actual, contingent behavior, and time varying data: reservoir recreation in the Columbia River basin. Working paper, Department of Agricultural Economics, University of Nevada, Reno.

Cameron, T.A. 1992. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Economics.* 68: 302-317.

Cole, R.A., F.A. Ward, T.J. Ward, and R.M. Wilson. 1990. Development of an Interdisciplinary Planning Model for Water and Fishery Management. *Water Resources Bulletin* 26 (4): 597-609.

Cordell, K. and J. Bergstrom. 1993. Comparison of recreation use values among alternative reservoir water level management scenarios. *Water Resources Research* 29, No. 2: 247-258.

Englin, J. and T.A. Cameron. 1993. "Comparing Observed and Multiple-Scenario Contingent Behavior: A Panel Analysis Utilizing Poisson Regression Techniques." *Sixth Interim Report of the W-133: Benefits & Costs Transfer in Natural Resource Planning.* pp. 266-278.

Florida Department of Environmental Protection (Economic Analysis Division). 1995. A Socioeconomic study of the Rodman Reservoir, Brad Bendel and Nick Stratis, Principle authors. Two volumes, Tallahassee, FA.

Heckman, J.J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic of Social Measurement.* 5(4): 475-492.

Hellerstein, D. 1992. Estimating consumer surplus in the censored linear model. *Land Economics*, 86: 83-92.

Kling, C.L. 1989. "A Note on the Welfare Effects of Omitting Substitute Prices and Qualities from Travel Costs Models." *Land Economics*, 65: 290-97.

Morey, E. R. 1981. The Demand for Site-Specific Recreational Activities: A Characteristics Approach. *J. of Environmental Economics and Management.* 8: 345-371.

Shaw, W.D. 1992. Searching for the Opportunity Cost of an Individual's Time. *Land Economics.* 68: 107-15.

Ward, F.A. 1989. Efficiently Managing Spatially Competing Water Uses: New Evidence from a Regional Recreation Demand Model. *J. of Regional Science.* 29 (2):229-46.

Table 1a

PROBIT MODELS FOR NON-RESPONSE SELECTIVITY-CORRECTION INVERSE MILLS RATIOS (MODELS FOR PRESENCE OF *ACTUAL* TRIP DATA)

VARIABLE NAME ones/zeros	REGION 1 n = 1428 547/881	REGION 2 n = 1432 513/919	REGION 3 n = 2092 746/1346	REGION 4 n = 1994 744/1250	VARIABLE DEFINITION
HAVDIS	0.19206 (0.65561)	5.7959 (3.1639)**	8.0259 (2.6931)**	9.9026 (1.9011)*	= 1 if distance data available; else 0
HAVOTH	0.24815 (0.78298)	-3.2686 (-1.4121)	-4.3770 (-1.3046)	-6.2987 (-1.1874)	= 1 if distances to other waters available; else 0
HAVCEN	0.50775 (1.3295)	-0.75287 (-1.6799)*	0.20172 (0.65698)	0.47260 (1.3083)	= 1 if census data available; else 0
P2	-0.56670 (-1.4236)	-0.60765E-01 (-0.18389)	-0.10002 (-0.36880)	-0.35417E-01 (-0.88460E-01)	= 1 if population 2; else 0 (adjacent counties)
P3	0.31880 (2.3157)**	0.24967 (1.9083)*	0.20273 (2.3221)**	0.32775 (3.0519)**	= 1 if population 3; else 0 (Phase 1A)
P4	0.17700 (1.2322)	0.38767 (1.9590)*	0.75925 (4.6284)**	1.4569 (2.4525)**	= 1 if population 4; else 0 (postcard sample)
P5	-0.77836 (-2.4479)**	-0.71648E-01 (-0.30487E-01)	0.20678 (0.70899E-01)	0.15327 (0.29853E-01)	= 1 if population 5; else 0 (Canada)
DIST1	0.76007 (3.8209)**	-0.15326E-01 (-0.47627E-01)	-0.22106 (-1.2452)	-0.75003 (-2.1654)**	distance to water 1
DIST2	0.15886E-01 (0.61097)	-0.75630 (-1.0439)	-0.15698 (-0.68196)	0.38345 (0.39464)	distance to water 2
DIST3	-2.3938 (-5.0610)**	0.39958 (0.60302)	-0.92492E-01 (-0.63125)	-0.75754 (-0.71607)	distance to water 3
DIST4	1.6753 (4.7493)**	0.16275 (1.3345)	-	0.11441 (0.28733)	distance to water 4
MIN1	1.3097 (2.6203)**	-0.59206 (-1.6082)	0.62133 (1.1583)	0.63910 (0.74619)	distance to nearest other water
MIN2	-0.46023 (-0.37122)	1.3676 (1.8142)*	-0.50887E-02 (-0.10216E-01)	-1.5657 (-1.3919)	distance to second nearest other water
MIN3	-3.5558 (-3.2208)**	0.55804 (0.99268)	-0.54104 (-0.70657)	1.6235 (2.2916)**	distance to third nearest other water
MIN4	-0.15599 (-0.16529)	-1.6106 (-2.4703)**	-0.43392E-01 (-0.11324)	-0.68595 (-0.79962)	distance to fourth nearest other water

MIN5	2.8234 (4.0153)**	0.52268 (0.88615)	0.48010 (1.2338)	0.86601 (1.7367)**	distance to fifth nearest other water
PURBAN	0.96628E-01 (0.50287)	-0.13891 (-0.72731)	0.14331 (1.0970)	0.27342 (1.9066)*	zipcode proportion urban
PBLACK	-8.9247 (-1.6234)*	1.6424 (0.24361)	-0.28036 (-0.42765)	-0.82496 (-1.0373)	zipcode proportion black
PAMIN	-1.4895 (-0.82022)	-4.2250 (-1.9716)**	4.0906 (1.7235)*	-1.0039 (-0.69557)	zipcode proportion American Indian
PASIAN	-1.1130 (-0.26646)	-4.6187 (-1.0410)	-6.8206 (-2.9792)**	0.25816 (0.18482)	zipcode proportion Asian
POTHER	-8.2914 (-2.2266)**	-1.5475 (-1.0175)	-1.7255 (-1.4537)	-0.95515 (-0.27191)	zipcode proportion other ethnicity
PLANGIS	-18.201 (-0.66426)	13.287 (0.86397)	25.447 (2.3780)**	-3.0029 (-0.27418)	zipcode proportion language-isolated
PCOLL	3.9241 (3.3720)**	0.76408 (0.81530)	0.73926 (1.4035)	0.47761 (0.79476)	zipcode proportion college grad and above
PAGIND	-43.012 (-3.7980)**	2.3255 (0.31369)	-2.0378 (-0.42363)	-2.1337 (-0.28822)	zipcode prop. in Ag, Forest, Fisheries industries
PAGOCC	45.638 (3.5426)**	-3.8934 (-0.44411)	0.62551 (0.10917)	0.35080 (0.46411E-01)	zipcode prop. in Ag, Forest, Fisheries occupations
PPUBINC	2.8946 (0.61562)	5.0396 (0.94489)	-3.3398 (-1.0000)	-9.8171 (-2.3600)**	zipcode proportion on Public Income Assistance
PSSINC	1.2595 (0.37603)	-2.0979 (-0.81532)	-2.1780 (-1.2694)	0.62706 (0.29149)	zipcode proportion on Social Security Income
PRETINC	-9.6987 (-1.8656)*	8.9402 (2.1858)**	1.3258 (0.53073)	2.2450 (0.76139)	zipcode proportion on Retirement Income
INCM	-0.20331E-01 (-1.6704)*	0.15675E-01 (2.2734)**	0.26583E-02 (0.48986)	-0.10602E-01 (-1.7994)*	zipcode median income
RENT	-0.21799 (-0.33908)	1.0826 (2.0317)**	0.20378 (0.54884)	-0.25236 (-0.53122)	zipcode median gross rental
VALUE	0.14280E-03 (0.63823E-01)	-0.89811E-02 (-3.6011)**	-0.32153E-02 (-1.8715)*	0.68629E-03 (0.61385)	zipcode median house value
CONSTANT	-0.42767 (-1.8358)*	-2.6230 (-1.1117)	-3.4912 (-1.3460)	-3.6669 (-0.73215)	intercept term
Log L	-850.93	-869.31	-1298.5	-1225.0	

NOTE: Distances DIST1, etc. refer to numbering of waters within each REGION and differ from numberings overall; different coefficients are allowed for each of the four different surveys, although the coefficients on the resulting inverse mills ratio terms will be constrained to be identical for each type of data.

Table 1b

PROBIT MODELS FOR NON-RESPONSE SELECTIVITY-CORRECTION INVERSE MILLS RATIOS (MODELS FOR PRESENCE OF *CONTINGENT* TRIP DATA)

VARIABLE NAME ones/zeros	REGION 1 n = 1428 388/1040	REGION 2 n = 1432 362/1070	REGION 3 n = 2092 480/1612	REGION 4 n = 1994 463/1531	VARIABLE DEFINITION
HAVDIS	-0.11081	3.7908	9.1345	3.8765	= 1 if distance data available; else 0
	(-0.35529)	(2.6406)**	(0.81051E-01)	(1.4132)	
HAVOTH	0.21367	-1.2880	-5.1020	-1.6757	= 1 if distances to other waters available; else 0
	(0.66918)	(-1.1958)	(-0.60290E-01)	(-0.81229)	
HAVCEN	0.55526	0.39420E-01	-0.37950E-02	0.33071	= 1 if census data available; else 0
	(1.2422)	(0.78558E-01)	(-0.11568E-01)	(0.83825)	
P2	-0.19400	0.22826E-01	0.47629E-01	0.35663E-01	= 1 if population 2; else 0 (adjacent counties)
	(-0.46454)	(0.66259E-01)	(0.16602)	(0.84455E-01)	
P3	0.15525	0.29310	0.99407E-01	0.12092	= 1 if population 3; else 0 (Phase 1A)
	(1.0163)	(2.1762)**	(1.0441)	(1.0416)	
P4	0.22938	0.59124	0.65270	0.67675	= 1 if population 4; else 0 (postcard sample)
	(1.4883)	(2.9344)**	(3.9663)**	(1.2054)	
P5	-0.50976	0.53861	-4.6319	-1.0760	= 1 if population 5; else 0 (Canada)
	(-1.5089)	(0.48933)	(-0.54737E-01)	(-1.5512)	
DIST1	0.56875	-0.37506	-0.25711	-0.49795	distance to water 1
	(2.8518)**	(-1.1294)	(-1.4806)	(-1.7122)*	
DIST2	0.11934	-0.33033	0.89893E-01	0.48583	distance to water 2
	(3.9794)**	(-0.44338)	(0.48033)	(0.45150)	
DIST3	-2.0713	0.32897	0.25313	-0.58558	distance to water 3
	(-4.5736)**	(0.49036)	(1.5799)	(-0.49977)	
DIST4	1.5746	0.21034	-	0.26921	distance to water 4
	(4.6357)**	(1.6949)*	(0.67650)		
MIN1	1.4648	-0.43656	-0.24237	-0.71304E-01	distance to nearest other water
	(2.6977)**	(-1.1298)	(-0.43911)	(-0.79093E-01)	
MIN2	-3.2871	1.4567	-0.44860	-0.28539	distance to second nearest other water
	(-2.4661)**	(1.7985)*	(-0.86074)	(-0.23702)	
MIN3	-0.52187	-0.28522	0.50755	0.80453	distance to third nearest other water
	(-0.44051)	(-0.48012)	(0.62903)	(1.0469)	
MIN4	0.25418	0.26563	-0.20090	-1.6494	distance to fourth nearest other water
	(0.24087)	(0.38213)	(-0.52612)	(-1.7662)*	

MIN5	1.9177	-0.78621	0.31533	1.1124	distance to fifth nearest other water
	(2.5336)**	(-1.2382)	(0.76164)	(2.0572)**	
PURBAN	0.13389	0.22053E-01	0.11820	0.11499	
	(0.59233)	(0.10965)	(0.83804)	(0.73126)	zipcode proportion urban
PBLACK	0.68181	5.4722	0.50597	-0.64461	
	(0.12518)	(0.82493)	(0.74346)	(-0.73452)	zipcode proportion black
PAMIN	-3.8575	-2.2895	-1.8527	-0.42595	
	(-1.8058)*	(-1.0894)	(-0.63621)	(-0.28104)	zipcode proportion American Indian
PASIAN	3.5603	-0.66362	-5.2765	-3.5888	
	(0.75824)	(-0.14070)	(-2.1121)**	(-2.2817)**	zipcode proportion Asian
POTHER	-2.6584	0.59555	-0.23971	-2.1420	
	(-0.57359)	(0.36316)	(-0.18928)	(-0.53483)	zipcode proportion other ethnicity
PLANGIS	-35.132	-17.117	11.359	19.195	
	(-1.1399)	(-1.0419)	(0.98728)	(1.6123)	zipcode proportion language-isolated
PCOLL	4.6255	-0.75101	0.69966	0.92313	
	(3.6276)**	(-0.75014)	(1.2199)	(1.4093)	zipcode proportion college grad and above
PAGIND	-63.915	-11.930	-3.3182	-0.38337	
	(-4.5474)**	(-1.5496)	(-0.63417)	(-0.46054E-01)	zipcode proportion employed Ag, Forest, Fisheries industr
PAGOCC	66.488	14.731	3.3218	-4.8356	
	(4.2532)**	(1.6126)	(0.53113)	(-0.56696)	zipcode proportion employed Ag, Forest, Fisheries occupat
PPUBINC	-6.2778	-1.0457	-2.0673	-8.6747	
	(-1.1771)	(-0.19043)	(-0.57221)	(-1.8256)*	zipcode proportion on Public Income Assistance
PSSINC	7.3902	-1.3841	-0.63432	-1.3483	
	(1.8655)*	(-0.51799)	(-0.33864)	(-0.56425)	zipcode proportion on Social Security Income
PRETINC	-10.514	6.5997	0.35290E-02	2.9735	
	(-1.7621)*	(1.5469)	(0.13092E-02)	(0.91239)	zipcode proportion on Retirement Income
INCM	-0.54540E-01	0.87124E-02	-0.14687E-02	-0.18455E-01	
	(-3.8116)**	(1.2214)	(-0.24497)	(-2.6070)**	zipcode median income
RENT	0.65800	0.72333	0.75581	0.73452	
	(0.91592)	(1.3420)	(1.8435)*	(1.3808)	zipcode median gross rental
VALUE	0.72889E-02	-0.45813E-02	-0.27965E-02	0.14200E-05	
	(2.9592)**	(-1.8108)*	(-1.4718)	(0.11163E-02)	zipcode median house value
CONSTANT	-0.96994	-3.3862	-5.1027	-2.5568	
	(-2.2919)**	(-3.9636)**	(-0.68556E-01)	(-2.3240)**	intercept term
Log L	-695.53	-759.24	-1063.6	-997.53	

NOTE: Distances DIST1, etc. refer to numbering of waters within each REGION and differ from numberings overall; different coefficients are allowed for each of the four different surveys, although the coefficients on the resulting inverse mills ratio terms will be constrained to be identical for each type of data.

TABLE 2
VARIABLES IN RECREATION DEMAND MODELS*

Variable Name	Variable Definition
PRICE	The own price of the water visit, equal to round trip distance calculated using the program ZIPFIP**, multiplied by the DOT estimate of 29 cents per mile for operating a vehicle, plus lodging costs, plus the opportunity cost of time in travel***
SFISH,BOAT,BFISH	These are the intercept shifter dummy variables: = 1 if the individual had a fishing license in 1993, owned a boat, had a fishing license and owned a boat.
RTVC1 - RTVC9	Cross price terms for each of the nine other projects
W1 through W9	The own water level for water X is reported as W_X for each of the models, the others from W1 - W9 are the potential cross project water levels
HAVDIS	= 1 if distance data were available for this individual; else 0
HAVINC,INC	= 1 if income reported for this individual; else 0, and annual income for 1993, if income data reported
OTHERA	Average price or distance for the other five closest waters ????
V1 through V4	Intercept shifter dummy for different survey versions when data are pooled
IMR2 and IMR3	Inverse mills ratios from the initial probit survey response models (revealed preference and stated preference response/non-response sample selection)
TOTTRIP	The total number of water-based recreation trips reported in each month for the Northeastern U.S. (controls for seasonal trip-taking behavior independent of historical water level management in Pacific Northwest)

* Models also include intercept terms, and dummy variables for whether the trips are taking place in the main summer months or during the remainder of the year.

** ZIPFIP calculated the road distance between two places using the latitude and longitude of the centroids for the respective zip codes. Comparison to the distances published in the AAA Road Atlas showed ZIPFIP estimates to be reasonably accurate.

*** Lodging costs are the sample average reported for each project by distance zone (< 25 miles, 26-149 miles, and > than 149 miles.) Opportunity cost of time is calculated for each individual by multiplying round trip distance divided by 40 mph assumed average speed, multiplied by the reported hourly wage rate (Shaw 1992).

Table 3a

First-Stage Probit and Second-Stage Tobit Parameter Estimates (Waters 1, 2, 3)

	Water 1 - Hungry Horse		Water 2 - Lake Pend Oreille		Water 3 - Lake Koocanusa	
	Probit (n = 1723)	Tobit (n = 399)	Probit (n = 4741)	Tobit (n = 648)	Probit (n = 1723)	Tobit (n = 419)
R_INT	0.89209 (3.103)**	1.2546 (1.085)	0.79977 (3.450)**	-1.2951 (-0.906)	0.79977 (3.450)**	-1.1202 (-1.255)
R_IMR2	-1143.0 (-0.003)	-1.5244 (-0.610)	-848.00 (-0.007)	9.2049 (1.609)	-848.00 (-0.007)	5.3601 (0.913)
PRICE	**-8.2493 (-5.098)****	**-38.633 (-2.432)****	**-13.107 (-5.771)****	**-18.339 (-2.422)****	**-13.107 (-5.771)****	**-34.351 (-2.549)****
INT	-30.656 (-5.590)**	-129.84 (-2.232)**	-100.05 (-1.170)	-100.09 (-0.203)	-100.05 (-1.170)	4.9706 (2.483)**
IMR2	-0.45949E-02 (-0.028)	-0.52688 (-1.195)	-0.34608 (-2.063)**	-3.1044 (-4.201)**	-0.34608 (-2.063)**	-2.0123 (-1.795)*
IMR3	-0.16408 (-0.807)	-0.94794 (-2.199)**	-0.28555 (-1.958)*	-2.4868 (-3.681)**	-0.28555 (-1.958)*	-0.57610 (-0.520)
V2						
V3				0.34247 (0.193)		
V4				0.95677 (0.945)		
HAVDIS	0.31101 (1.109)	1.3196 (1.513)	-0.27614 (-1.188)	0.63147 (0.130)	-0.27614 (-1.188)	-2.6869 (-2.972)**
RTVC1			8.1113 (6.177)**	-10.095 (-0.858)	8.1113 (6.177)**	14.796 (1.773)*
RTVC2	-6.7632 (-3.565)**	-31.694 (-2.190)**			7.9179 (6.227)**	-0.56485 (-0.057)
RTVC3	-1.9091 (-0.658)	-8.0547 (-0.749)	7.9179 (6.227)**	31.011 (1.118)		
RTVC4	14.966 (4.328)**	68.472 (2.145)**	-2.8495 (-1.418)	-11.127 (-0.576)	-2.8495 (-1.418)	17.495 (1.309)
RTVC5				10.771 (0.945)		
RTVC6				8.0158 (1.864)*		
RTVC7				-8.5776 (-0.708)		
RTVC8				10.453 (1.880)*		
RTVC9						

OTHERA	0.46007 (0.431)	2.3306 (0.303)	-0.89741 (-0.641)	13.103 (1.267)	-0.89741 (-0.641)	3.5126 (0.433)
HAVINC	-0.14441 (-0.764)	-0.15744 (-0.249)	-0.24670 (-1.755)*	-1.7673 (-2.856)**	-0.24670 (-1.755)*	0.52155E-01 (0.073)
INC	3.1031 (1.332)	6.5722 (0.742)	-1.2527 (-0.666)	0.70318 (0.064)	-1.2527 (-0.666)	25.576 (3.258)**
SFISH	0.40769 (2.740)**	1.7474 (2.552)**	0.35034 (2.568)**		0.35034 (2.568)**	0.94986 (1.991)**
BOAT	0.40394E-01 (0.341)	0.31488 (1.448)	0.51254 (4.478)**		0.51254 (4.478)**	1.8696 (2.476)**
BFISH	-0.61982 (-3.103)**	-2.3064 (-2.239)**	-0.41468 (-2.169)**		-0.41468 (-2.169)**	
W1S	**5.9381** **(4.408)****	**26.927** **(2.561)****				
W2S			**44.598** **(1.047)**	**51.514** **(0.217)**	44.598 (1.047)	
W3S	3.7086 (2.968)**	12.439 (1.608)	3.2562 (2.351)**		**3.2562** **(2.351)****	
W4S						
W5S						
W6S						
W7S						
W8S						
W9S						
TOTTRIP	0.14071 (1.946)*	0.56754 (1.817)*	0.21575E-01 (0.233)	-0.13249 (-0.432)	0.21575E-01 (0.233)	-0.16679E-01 (-0.058)
λ_M		4.8651 (2.064)**		0.38699 (0.175)		-0.26818 (-0.190)
λ_R		0.96481E-01 (1.410)		-0.98536 (-0.431)		-0.20597E-01 (-0.127)
λ_A		6.0796 (2.016)**		0.72960 (0.288)		-1.1411 (-0.495)
RST	5.0704 (0.016)	-0.89447 (-4.347)**	4.7825 (0.034)	-0.67981 (-4.145)**	4.7825 (0.034)	-0.82107 (-2.885)**
ANN	0.38654E-01 (0.288)	-0.21890 (-3.999)**	-0.27129 (-1.854)*	-0.35862 (-8.257)**	-0.27129 (-1.854)*	-0.17666E-01 (-0.322)
σ		1.4172 (21.206)**		3.7088 (30.856)**		2.8262 (20.988)**
log L	-685.64	-643.87	-1423.0	-1625.6	-729.08	-1012.8

Table 3b

First-Stage Probit and Second-Stage Tobit Parameter Estimates (Waters 4, 5, and 6)

	Water 4 - Kootenai River		Water 5 - Dworshak Lake		Water 6 - Clearwater River	
	Probit (n = 1723)	Tobit (n = 363)	Probit (n = 3018)	Tobit (n = 555)	Probit (n = 1512)	Tobit (n = 319)
R_INT	0.60261 (2.915)**	-2.1347 (-1.156)	1.1646 (3.110)**	8.3581 (1.978)**	1.7831 (4.727)**	3.2082 (0.796)
R_IMR2	-848.00 (-0.007)	-0.11522 (-0.011)	-519.40 (-0.019)	-19.203 (-0.964)	-132.00 (-0.047)	-16.210 (-2.320)**
R_V4			-0.10556 (-0.524)	-0.80593 (-0.250)		
PRICE	**-5.7554 (-2.871)****	**-50.825 (-2.605)****	**-7.5627 (-9.686)****	**-55.223 (-3.461)****	**-5.7628 (-0.799)**	**-94.506 (-3.067)****
INT	-11.689 (-3.549)**	9.7269 (3.311)**	-5.3763 (-0.872)	-55.900 (-2.413)**	14.374 (1.997)**	-12.757 (-0.975)
IMR2	0.47487E-01 (0.333)	-4.1379 (-5.412)**	-1.4267 (-6.352)**	-10.429 (-2.919)**	0.31135 (1.796)*	2.2208 (2.058)**
IMR3	0.21493 (1.506)	-2.6098 (-3.514)**	-1.4458 (-5.245)**	-11.404 (-1.911)*	-0.47136 (-1.483)	-1.5899 (-1.302)
V4			-0.48671 (-2.774)**	-4.4965 (-1.532)		
HAVDIS	-0.48479 (-2.190)**	0.71430E-01 (0.029)	-0.77063 (-1.464)	-3.6462 (-0.517)	-1.5239 (-2.170)**	-4.2979 (-0.478)
RTVC1	7.0574 (6.038)**	-24.947 (-1.957)*				
RTVC2	6.6666 (4.853)**	1.0345 (0.063)	1.1547 (2.969)**	6.3733 (1.272)	2.3214 (4.117)**	14.652 (2.448)**
RTVC3	-9.4152 (-4.561)**	73.516 (3.263)**				
RTVC4						
RTVC5					0.25692 (0.129)	13.692 (1.608)
RTVC6			1.3603 (2.559)**	9.3315 (0.654)		
RTVC7			2.5207 (3.236)**	12.075 (1.287)	0.96137 (0.148)	63.676 (2.576)**
RTVC8			2.9130 (5.452)**	21.249 (2.184)**		
RTVC9						

OTHERA	2.8313 (2.693)**	4.4340 (0.208)	1.4904 (1.021)	28.883 (1.187)	1.1441 (0.667)	9.4333 (1.188)
HAVINC	-0.27216 (-2.073)**	2.5387 (1.929)*	-0.16836 (-1.121)	-0.78984 (-0.488)	0.27536 (1.277)	2.9458 (2.716)**
INC	6.4135 (3.581)**	-38.935 (-2.295)**	-0.16046 (-0.090)	6.0731 (0.266)	-5.1892 (-1.635)	-55.160 (-3.246)**
SFISH	0.41153 (3.293)**		0.42942 (2.796)**	3.4594 (2.901)**		-0.53577 (-1.102)
BOAT	0.16393 (1.784)*		0.55258 (5.143)**	4.7866 (2.050)**		-0.82709 (-1.977)**
BFISH	-0.15198 (-0.847)		-0.45767 (-2.299)**	-4.1940 (-1.747)*		1.3783 (1.892)*
W1S						
W2S						
W3S	4.4824 (3.249)**					
W4S	**-0.41448E-01 (-1.894)***	**0.28617E-01 (0.327)**				
W5S			**3.8500 (1.464)**	**35.320 (2.486)****	-10.773 (-2.645)**	
W6S			0.26173 (1.391)		**0.19447 (0.954)**	**0.74587 (1.111)**
W7S			-3.0270 (-0.550)			
W8S						
W9S						
TOTTRIP	0.17643 (2.135)**	-0.94855E-01 (-0.205)	0.15296 (2.075)**	0.76509 (2.035)**	0.28512 (3.108)**	0.69065 (0.839)
λ_M		-3.8545 (-2.143)**		9.3198 (2.758)**		3.6077 (1.116)
λ_R		0.28958 (0.989)		0.37805 (1.050)		0.53991 (2.670)**
λ_A		-5.9967 (-2.247)**		9.9007 (1.672)*		5.8493 (1.509)
RST	4.8533 (0.034)	-0.79438 (-3.745)**	4.2193 (0.081)	0.47994E-01 (0.316)	3.3909 (0.157)	-0.66334 (-5.144)**
ANN	-0.35914 (-2.585)**	-0.18132 (-3.225)**	0.61543E-01 (0.708)	0.83383 (19.293)**	0.31397 (1.574)	-0.47685 (-5.930)**
σ		3.8905 (22.238)**		2.6627 (28.243)**		3.2445 (15.523)**
Log L	-681.22	-962.07	-865.56	-1554.7	-622.68	-720.40

Table 3c

First-Stage Probit and Second-Stage Tobit Parameter Estimates (Waters 7, 8 and 9)

	Water 7 - Lower Granite Lake		Water 8 - Lake Roosevelt		Water 9 - Lake Umatilla	
	Probit (n = 4642)	Tobit (n = 573)	Probit (n = 3130)	Tobit (n = 696)	Probit (n = 1615)	Tobit (n = 281)
R_INT	-1.1518 (-0.941)	-2.1428 (-1.344)	0.41173 (1.330)	-0.49542 (-0.524)	-0.13577E-01 (-0.070)	-1.0735 (-0.852)
R_IMR2	-0.25969 (-0.488)	-12.682 (-2.634)**	-1.3807 (-0.970)	10.278 (0.591)	-0.13980E-01 (-0.070)	-0.16894 (-0.028)
R_V3	-0.92697E-01 (-0.858)	-0.17261E-01 (-0.024)				
R_V4	-0.52230E-01 (-0.747)	0.18698 (0.306)	0.14252 (1.231)	-0.26951 (-0.493)		
PRICE	**-4.3846 (-12.117)****	**-6.7390 (-1.018)**	**-6.3528 (-10.259)****	**-2.9162 (-0.350)**	**-1.3500 (-3.386)****	**-3.2134 (-0.944)**
PSFISH			0.77040 (1.306)	0.93980 (0.345)		-3.6012 (-1.669)*
PBOAT			0.65047 (1.265)	1.1744 (0.413)		
PBFISH			-0.50705E-01 (-0.061)	0.71399 (0.203)		
PIMR2						0.23989E-01 (0.007)
PIMR3						0.17492E-01 (0.007)
INT	-152.33 (-0.107)	-9.6803 (-0.499)	-157.90 (-3.081)**	-18.744 (-1.058)	-4.4900 (-1.000)	-20.943 (-1.376)
IMR2	-0.10581 (-0.862)	-1.5355 (-1.575)	-0.20532 (-1.340)	1.6207 (2.439)**	-1.0900 (-7.851)**	-4.3542 (-1.620)
IMR3	-0.10638 (-0.653)	-0.78093 (-1.217)	-0.74693E-04 (0.000)	1.1635 (2.522)**	-0.30800 (-1.575)	-1.0479 (-1.178)
V3	-0.94409 (-6.278)**	-2.0171 (-1.049)				
V4	-0.26197 (-2.043)**	-2.0959 (-2.044)**	0.21642 (1.577)	0.54974 (0.789)		
CONT	1.8724 (1.507)					
HAVDIS	3.9134 (0.003)		-0.44868 (-1.359)	10.118 (0.824)		
RTVC1						
RTVC2	0.18164E-01 (0.037)	3.1505 (0.889)	1.5391 (2.784)**	-0.40370 (-0.103)		
RTVC3						

RTVC4						
RTVC5	-0.48471 (-0.956)	-0.44534 (-0.094)	-1.9885 (-3.092)**	0.87737 (0.192)		
RTVC6	-1.7313 (-3.385)**	-7.7535 (-1.262)				
RTVC7			3.8053 (8.287)**	-0.45231 (-0.091)	-1.6600 (-3.668)**	-9.5295 (-2.250)**
RTVC8	2.6868 (7.380)**	5.8555 (1.468)			2.4200 (5.015)**	11.718 (1.782)*
RTVC9	1.4086 (3.584)**	1.1080 (0.251)	0.50457E-01 (0.105)	-0.17224 (-0.085)		
OTHERA	-0.21643E-01 (-0.015)	-5.2554 (-0.332)	-0.52264 (-0.508)	-7.7420 (-1.184)	1.8100 (1.002)	
HAVINC	-0.14686E-01 (-0.115)	-0.33528 (-0.604)	0.15541 (1.181)	0.37571 (0.728)	-0.45000 (-2.231)**	-2.0840 (-2.192)**
INC	0.85214 (0.433)	-7.1893 (-0.779)	1.6879 (0.816)	5.0650 (0.550)	1.1100 (0.372)	2.7488 (0.296)
SFISH	-0.13288 (-1.363)		0.21932 (1.397)		0.53793E-02 (0.070)	-0.78873 (-2.328)**
BOAT	0.77277E-02 (0.093)		0.93401E-01 (0.616)			1.3459 (2.241)**
BFISH	0.21745 (1.635)		-0.25595 (-1.107)			1.0390 (1.490)
W1S						
W2S	66.783 (2.568)**		72.223 (2.891)**			
W3S						
W4S						
W5S						
W6S						
W7S	13.816 (3.401)**	19.641 (0.826)				
W8S			6.9832 (2.204)**	7.5170 (0.700)		
W9S					17.000 (0.999)	71.178 (1.396)

TOTTRIP	0.36580E-01 (0.616)	0.30533E-01 (0.109)	-0.41031E-02 (-0.068)	0.12750E-01 (0.072)	0.37098E-01 (0.522)	0.40194 (2.271)**
λ_M		0.34674 (0.181)		-0.22196 (-0.148)		5.0123 (1.675)*
λ_R		4.6936 (4.953)**		-9.0818 (-0.712)		0.89259 (0.702)
λ_A		1.1673 (0.515)		-0.37916 (-0.199)		4.9743 (1.432)
RST	-1.4453 (-1.249)	-0.72367 (-4.399)**	-0.87759 (-0.964)	-0.23996 (-1.559)	-5.4100 (-0.377)	-0.32595 (-1.194)
ANN	0.77558E-01 (1.017)	-0.42761 (-10.700)**	-0.18128 (-1.897)*	-0.17881 (-3.885)**	-0.27399E-01 (-0.140)	0.15747 (2.157)**
σ		3.1241 (29.706)**		2.2127 (30.656)**		1.2177 (18.201)**
Log L	-1334.8	-1325.3	-1224.8	-1477.1	-663.54	-468.64

Rural Household Recycling: Explaining Participation and Weight Generation

by

Paul M. Jakus, Kelly H. Tiller and William M. Park

P.O. Box 1071

Dept. of Agricultural Economics and Rural Sociology

University of Tennessee

Knoxville, TN 37901-1071

Rural Household Recycling: Explaining Participation and Weight Generation

Abstract

Rising landfill costs have forced solid waste managers to consider different means to reduce the waste stream. We present a theory of household recycling in which recycling occurs in the absence of monetary incentives. This is important in rural regions of the country where unit-based garbage disposal fees are not an option. Results indicate that households are responsive to the own-cost effects of recycling. Waste managers may influence recycling participation and weight generation through programs designed to lower perceived time costs of non-recyclers, and improve the efficiency of recyclers.

Rural Household Recycling: Explaining Participation and Weight Generation

I. Introduction

Rural counties and small communities with low population densities throughout the United States face new constraints and pressures with regard to solid waste management. Environmental concerns (such as groundwater quality) have led to stricter landfill regulations, increasing the cost of traditional solid waste disposal methods (Darcey, 1991). Some states and local governments have mandated reductions in the amount of solid waste requiring disposal and/or have required recycling programs. Demands by citizens for recycling opportunities further intensify these pressures, forcing local decision makers to consider a wide range of alternative solid waste management plans (SWMP). For example, the Tennessee Solid Waste Management Act of 1991 places great emphasis upon waste reduction and recycling, mandating a 25% reduction in the per capita weight of solid waste burned or buried over a six-year period. Each county is also required to provide at least one site for collection of recyclable material, as well as a system of "convenience centers" at which rural residents may dispose of garbage. Because of the significant economies of size in collection, processing, and marketing of recyclables, a large volume of materials must be recovered if residential recycling is to be a cost-effective component of a SWMP. Such economies are difficult to achieve in rural areas unless participation rates are high and a large proportion of eligible material is recycled.

Although information regarding potential recycling participation rates and generation is crucial in determining whether a recycling program can be an efficient component of a SWMP, there has been little research investigating the economic factors which influence household recycling participation rates and generation. The few available studies focus on curbside collection of recyclables in urban areas. These studies have examined the effect on the quantity of material recycled in response to unit-based pricing for garbage disposal (Hong, et al, 1993; Morris and Holthausen, 1994; Fullerton and Kinnaman, 1993). There has been no attempt to measure an own-price elasticity for recycling.

In contrast, this study examines household recycling decisions in rural areas where low population density renders curbside recycling impractical. Recycling does not take place at the curbside; rather, recyclers must transport recyclables to a dropoff site. The factors influencing household recycling participation decisions are identified and the elements of a rural recycling program designed to encourage household recycling are evaluated. While options available to encourage recycling participation appear limited, the weight of material recycled is found to be responsive to the time cost of recycling. An estimate of the own-price elasticity of demand for recycling is reported.

This research provides solid waste managers with information useful in the design of recycling programs in rural areas where curbside recycling is not an economically viable option. Further, given the relatively high cost of curbside collection (up to $200/ton), some urban communities may consider changing from curbside collection to a system of dropoff centers which can significantly lower costs imposed on public agencies (with collection costs of less than $100/ton). Because dropoff programs shift some recycling costs to consumers, it is important to gauge how consumers will respond. This study provides an estimate of average household production of recyclables within the context of a dropoff system, which may prove useful in comparison with estimates from existing curbside collection systems.

II. A Behavioral Model of Recycling

Theoretical Model

In this section, a model is presented which includes motivations to recycle beyond avoiding increased garbage collection fees. Recycling in the absence of monetary incentives is observed in numerous communities; it is often observed even when costly in terms of time and effort. The model thus explicitly recognizes that some individuals may wish to limit the amount of waste generated and sent to a landfill or incinerator. The utility function is given by,

$$U[\, Z(x),\, G(S,x),\, L]$$

Z is the consumption commodity produced by the household using inputs **x**, where **x** is a (n x 1) vector of marketed goods. G is the amount of garbage sent for disposal, and is a function of inputs **x** and time spent separating recyclables, **S**. **S** is a (n x 1) vector of time requirements to recycle some portion of the refuse generated by marketed goods **x**. L is the amount of leisure consumed. The marginal product (Z_{xj}) of any element j in **x** is positive, while the marginal utilities are assumed to be $U_Z>0$, $U_L>0$, and $U_G \leq 0$. This last term is an inequality since garbage generation will impact the utility of some people negatively (those who would consider voluntary recycling), while it will not affect others (those who do not care about waste production).[1]

Use of inputs **x** generates trash, T, according to a function T(**x**), where $T_{xj}>0$. Trash may be separated into garbage or recyclable materials. Production of recyclables, R, is a function of the time spent in separation of the recyclables, **S**, and the commodity, **x**,

$$R = R(\mathbf{S}, \mathbf{x}) \tag{1}$$

where R is increasing in both arguments. The amount of garbage is determined by T-R, or

$$G(\mathbf{S}, \mathbf{x}) = T(\mathbf{x}) - R(\mathbf{S}, \mathbf{x}) \tag{2}$$

Let the household's full income consist of wage and non-wage income, so the budget constraint is,

$$wH + V = \mathbf{p'x} + fG(\mathbf{S}, \mathbf{x})$$

where w is the wage rate, H is hours worked, V is non-labor income, f is the unit cost of garbage disposal, and **p** is the (n x 1) price vector for **x**. The household's time is also constrained according to,

$$T = H + L + \mathbf{i'S}$$

where T is total time available and **i** is a (n x 1) vector of ones. Substituting (1) and (2) directly into the utility function and budget constraint yields a consumer problem in which the variables of interest are **x**, **S**, and L. The constrained optimization problem is then given by (3),

$$\max \mathcal{L} = U[Z(\mathbf{x}), T(\mathbf{x})\text{-}R(\mathbf{S},\mathbf{x}), L] + \lambda\{wH + V - \mathbf{p'x} - f[T(\mathbf{x})\text{-}R(\mathbf{S},\mathbf{x})]\}$$
$$+ \mu\{T - H - L - \mathbf{i'S}\} \tag{3}$$

The conditions needed to optimize **x**, **S**, and L are given by (4a) through (4f),

$$\partial\mathcal{L}/\partial x_j = U_Z Z_{xj} + U_G[T_{xj} - R_{xj}] - \lambda[p_j + f(T_{xj} - R_{xj})] \leq 0 \qquad (4a)$$

$$j = 1,\ldots, n$$

$$[\partial\mathcal{L}/\partial x_j]x_j = 0 \qquad (4b)$$

$$\partial\mathcal{L}/\partial S_j = - U_G R_{Sj} + \lambda[fR_{Sj}] - \mu \leq 0, \qquad (4c)$$

$$j = 1,\ldots, n$$

$$[\partial\mathcal{L}/\partial S_j]S_j = 0 \qquad (4d)$$

$$\partial\mathcal{L}/\partial L = U_L - \mu \leq 0, \qquad (4e)$$

$$[\partial\mathcal{L}/\partial L]L = 0 \qquad (4f)$$

where λ is the shadow value of income, μ is the shadow value of time, and Kuhn-Tucker conditions are needed since not all consumers will choose to recycle.[2]

Equation (4a) shows that the choice of optimal x_j is affected by the "marginal utility product" of x_j and the potential disutility of garbage produced (if $U_G<0$), where $(T_{xj}-R_{xj})$ is the amount of garbage generated by the marginal unit of x_j. Condition (4a) reflects not only the market price of x_j, but also its disposal cost, $f(T_{xj}-R_{xj})$. Condition (4b) is presented for completeness.

Equations (4c) and (4d) govern the optimal choice of S_j, the amount of time invested in materials preparation and separation for good x_j. This condition is best discussed by considering a consumer who chooses to recycle. When (4c) is an equality, division by λ monetizes all terms, so that the marginal benefit of time spent recycling is just balanced by the net marginal cost of recycling, $[\mu/\lambda - fR_{Sj}]$, where μ/λ is the opportunity cost of time. If an individual's marginal cost of recycling exceeds the marginal benefit (e.g., if $U_G=0$ and $\mu/\lambda>fR_{Sj}$), then (4c) is negative and (4d) represents the appropriate marginal condition for a consumer operating at a corner solution (not recycling).

Conditions (4e) and (4f) deal with the optimal choice of leisure. At an interior solution, the marginal utility of leisure is equated with the shadow value of time. If the leisure activity is the next best alternative to recycling activities (if, for example, work hours H cannot be adjusted by the individual), the opportunity cost of time in (4c) is not necessarily equal to the wage rate.[3]

The model explains how people can engage in a variety of waste reduction and recycling activities. Upon an increase in the current disposal fee (or imposition of a unit-based fee), consumers may choose products which limit disposable packaging, i.e., they choose products to decrease T(**x**) or increase R(**S**, **x**). Because f enters the budget constraint, consumers will "waste reduce" even if $U_G=0$ [by condition (4a)]. Second, some consumers not currently recycling - those with $U_G<0$ but whose marginal costs exceed the marginal benefits - would be more likely to recycle as the marginal cost of recycling declines [by condition (4c)]. Finally, the model provides a rationale for the observation that some people recycle even in the absence of monetary incentives. Flat-fee garbage disposal prices would cause marginal disposal cost to drop out of the first order conditions. Although more consumption of x_j would be observed as x_j becomes relatively less expensive, and less time spent in separation would be observed as S_j becomes relatively more expensive, the behavioral motivation to recycle remains [by condition (4c)].

Implementing the Model

Even simple household production models require restrictive assumptions regarding the technology available to the household to make such models empirically tractable. Inputs to and outputs from each production process must be precisely identifiable and non-joint. Production processes must have linear cost functions to yield a linear budget constraint. Finally, the technical coefficients of the cost function must be exogenous to the consumer (Pollack and Wachter, 1975). As written, however, the model exhibits a high degree of jointness in the marketed goods vector **x**, where **x** is an input to the consumption, waste generation, and recycling production processes. Fortunately, some reasonable assumptions regarding the waste generation and recycling production technologies are sufficient to meet the conditions required for empirical estimation of the model.

Begin by noting that each marketed commodity can be separated by composition: product and packaging. For example, a soft drink may contain 16 ounces of product (its net weight) and 1 ounce of plastic packaging. x_j is then allocable across the production technologies, such that any unit of good x_j contains θ_j proportion of packaging and $(1-\theta_j)$ of product. The packaging coefficient θ_j is exogenous and bounded by zero and one. Each good x_j has its own packaging coefficient, so that total trash production is $T(\mathbf{x}) = \sum_j \theta_j x_j$. Consumers may engage in waste reduction through judicious selection of marketed goods with respect to relative amounts of packaging and product. The allocation of **x** establishes separability of Z(**x**) and G(**S**, **x**), which may be re-written as $Z[(i-\theta)'\mathbf{x}]$ and $G(\mathbf{S}, \theta'\mathbf{x})$, where θ is a (n x 1) packaging coefficient vector.[4]

Each unit of trash can also be divided into its recyclable and non-recyclable components. For refuse generated by input x_j, let γ_j be the proportion which is recyclable and $(1-\gamma_j)$ be the proportion which is not recyclable, where γ_j is exogenous and bounded by zero and one. Further, the recyclable material $(\gamma_j\theta_j x_j)$ also requires τ_j units of time spent recycling refuse from x_j. Where S_j is time spent recycling x_j, the production function for recyclables may be written as,

$$r_j = \min\,[\tau_j S_j,\ \gamma_j(\theta_j x_j)]$$

Assuming that x_j is purchased for its product rather than its packaging, packaging available for disposal or recycling is costless. Noting that recycling requires little or no costs other than time, the unit cost function associated with recyclables production is,

$$c_j = (\omega/\tau_j) \tag{5}$$

where ω is equal to the opportunity cost of time, μ/λ.[5] The τ_j and γ_j are exogenous to the consumer and the cost function is linear, satisfying the conditions outlined by Pollack and Wachter.

With these restrictions, solution to (3) yields a household generation function for recyclable material. The function measures the aggregate weight of recyclable material generated by the household, and has as its arguments all cost and income terms,

$$R = R(\omega/\tau_1,\ldots,\omega/\tau_n,\ \theta,\ \gamma,\ f,\ \mathbf{p},\ wH+V) \tag{6}$$

R(.) is expected to be negatively influenced by the recycling cost and positively related to garbage disposal costs and income. It is not clear *a priori* how R(.) will be impacted by **p** because changes in relative prices will result in substitution between products with different packaging and recycling coefficients.

III. Data

Household recycling data were collected in Williamson County, Tennessee in August and November of 1992. Located in middle Tennessee just south of Nashville, Williamson County is the state's most affluent county (as measured by household median income). The county has distinct suburban (northern) and rural (southern) portions. Most households in rural areas do not contract for house-to-house garbage collection. It is illegal to burn or bury waste, so the county has established a network of seven convenience (dropoff) centers in rural areas, where residents without house-to-house garbage collection can drop off their garbage. Residents are not required to separate trash into garbage and recyclables, but separated recyclables are accepted at all convenience centers. The vast majority of rural residents live less than 5 miles from the nearest convenience center, so travel costs are negligible.[6]

In addition to convenience centers, Williamson County's recycling program includes a material recovery facility for intermediate processing and a full-time program coordinator. The program coordinator disseminates information to the public and oversees a recycling education program in elementary and middle-schools. Revenue from the sale of the recycled material from each convenience center is donated to non-profit groups located in the vicinity of the convenience center.

The survey instrument was designed using a focus group and two pre-tests conducted at convenience centers in Knox County, Tennessee. Two hundred eighty-four individuals were interviewed as they entered convenience centers and asked if they would participate in the study.[7] Upon completion of one interview, enumerators attempted to interview the next person entering the convenience center. The refusal rate was 29.85%. Table 1 presents a statistical profile of the respondents. Respondents were presented with a number of statements regarding issues associated with household recycling and rural solid waste management, and were asked to state the degree to which they agreed or disagreed with these statements. The statements and mean response are also reported in Table 1.

As each respondent was being questioned by one interviewer, his or her garbage and/or recyclables were being weighed by a second interviewer. Respondents were asked to estimate the number of days it had taken to accumulate the garbage and/or recyclables they had brought with them. Accumulation intervals were used to convert measured weights to monthly generation rates for each of the five recyclable materials accepted in the recycling program.[8] Table 2 reports recycling participation rates and generation rates.

It is useful to note the large differences between self-reported recycling participation and respondents who actually brought recyclables on the day of the interview. Overall, 75% of respondents stated they recycled at least one material, but only 31% actually brought any recyclables on the day of the interview. One would not expect all recyclers to bring all recyclables each time they bring garbage for dropoff; however, the disparity between the percentages is striking and requires some explanation. For many people the accumulation interval for recyclables is longer than that for garbage. This may be because recyclables "keep" better than garbage, or that storage capacity and household size combine to determine frequency of recyclables dropoff. An in-person survey method is more likely to intercept a recycler on a "garbage-only day" rather than a day on which they also brought recyclables. Further, a household with relatively high recycling costs would accumulate recyclables at a slower rate than those with lower costs.[9] This issue is addressed in discussion of appropriate econometric methods.

IV. Empirical Models

This section reports models of household participation and volume generation estimated using two-stage methods. For recycling participation, a grouped data model predicting household income is presented as a first-stage model. The second-stage models estimate the probability (for any material) that equation (4c) holds as an equality. For recyclables weight generation, first-stage cost function models corresponding to equation (5) in Section II are estimated. Second-stage generation models follow, which correspond to equation (6). Finally, the results of the participation and generation models are examined with respect to their implications for the design of rural recycling programs.

Participation Models

First-Stage: An Income Model

Survey respondents may be sensitive to some questions and choose not to respond, especially open-ended questions about household income. To minimize non-response, researchers often ask respondents to indicate a range within which their income falls, but this leaves a discrete, rather than continuous, indicator of income. Because these data provide only upper and lower bounds on income, analysts often employ the midpoint of each range in estimating their models, with a sometimes arbitrary decision made in assigning a midpoint to the open-ended upper range. A useful alternative to this ad hoc method is a grouped data model (Stewart, 1983). Such a model adjusts for the doubly censored nature of the discrete data, converting the discrete variable into a continuous variable predicting income as a function of consumer characteristics. Table 3 presents the model used in this analysis, where all variables are consistent with expectations. Predicted income (INCOME) will be used as an explanatory variable in both participation models and generation models. The model predicted negative incomes for 12 observations, which were dropped from subsequent analysis.

Second Stage: Recycling Participation

Participation models are presented in Table 4. These models use probit analysis to gauge the factors determining whether equation (4c) holds as an equality (recycler) or an inequality (non-recycler). The explanatory variables are designated by two categories: those variables capturing the influence of household production technology and characteristics, and those variables capturing the impact of the Williamson County recycling program. In general, the models are consistent with theoretical expectations, with nearly all variables having the expected sign and most highly significant.

Specification #1 includes all variables believed to influence the decision to recycle. Some household production and characteristics variables were coded 1, 2, 3 or 4 such that low codes reflected "recycling" behavior and higher codes reflected "non-recycling" behavior (see Table 1 for variable definitions). Thus, the expected sign on these variables was negative. Households believing that recycling takes little time were more likely to recycle than those disagreeing with this statement (TIME). This represents a rough measure of household's perception of τ_j where, all else equal, a low τ_j suggests a greater marginal return per unit of time spent recycling. Other factors could also affect the decision to recycle, particularly those which may constrain the elements of the marginal benefit term in (4c). Households believing that they generate enough material to warrant recycling were more likely to recycle (GENERATE). Further, households with adequate storage space were also more likely to recycle

(STORAGE). Finally, other household characteristics may influence behavior as well. Households were more likely to recycle if they had friends who recycled (FRIENDS) or as the respondent's age increased (AGE).[10] A respondent's education (EDUCATION) and the household's income (INCOME) were not significant determinants in the recycling decision.[11]

Specification #1 also included variables designed to capture the impact of the Williamson County information and education program. This program is explicitly designed to influence residents' sensitivity to solid waste issues. CHILDREN 6-14 attempts to capture the influence of the elementary and middle-school education program, while DONATE measures whether the respondent was aware that revenue went to local non-profit organizations. INFORMATION measured the overall effectiveness of the program in educating the public about recycling and solid waste issues. Individually, none of the program specific variables were statistically significant. The hypothesis that the slopes of the program variables were jointly equal to zero was not rejected (model #1 vs. model #2, $\chi^2=1.64$). Specification #3 estimated participation as a function of only the program variables, none of which were significant at $\alpha=.05$, while the entire model was not significant at conventional levels ($p=0.12$). The hypothesis that the household production technology and characteristics slopes were jointly equal to zero was soundly rejected (model #1 vs. model #3, $\chi^2=58.01$).

Before concluding that the recycling program is ineffective, however, two caveats must be stated. First, the variables used to gauge the program are crude measures. In particular, CHILDREN 6-14 may capture not only the impact of the school program, but also other factors associated with having school-aged children in the household. A better measure would differentiate between those households with school-aged children and those whose children had participated in the recycling education program. Second, the effectiveness of the recycling education program may best be evaluated by observing how the marginal effects change over time. For example, while having school-aged children in the household may limit recycling participation (i.e., give a negative sign in a probit participation model), an school recycling program may make this effect "less negative". If households are more likely to recycle after a school program than they were before, the program is successful even if the probit model yields a negative coefficient.

Generation Models

Only those people saying they recycle and actually bringing recyclable material on the day of the interview exhibit non-zero generation. The data thus contain two kinds of "zero-generation" observations: those who do not recycle, and those who do recycle but brought no material to the dropoff center on the day of the interview. Pudney (1989, p. 174) refers to this second kind of zero as a "fortuitous" observation, where a household may consume goods (produce recyclables) in the long-run, but is surveyed during a time period in which no purchases are made. Fortuitous observations should be retained since there exist valid reasons for observing zero generation. In this sample, each household representative who claimed to recycle a material was asked how long, on average, it took to recycle one unit of the material (i.e., one newspaper, a glass bottle, etc.). For all materials, the mean recycling time requirement reported by those who actually brought material that day was less than the mean time requirement reported by those who said they recycled the material, but did not bring any. Further, individuals with longer accumulation intervals tended to give higher estimates of recycling time. These results are consistent with the hypothesis that those households with zero generation on a given day experience higher costs or are less efficient recyclers, and are therefore less likely to generate as great a weight as more efficient recyclers. Thus, two selection criteria are in operation: first, does the household recycle this material, and second, was any material brought to the dropoff center? Catsiapis and Robinson (1978) were the first to empirically address this type of problem using OLS, with Lee (1983) providing a general econometric model. A probit selection model is used for the recycle/not recycle decision, estimated simultaneously with a tobit model for the generation models, where the dependent variable for recyclers (weight of recyclables) is censored at zero (Greene, 1991, p.572).[12]

<u>First Stage: Cost Functions</u>

The participation models indicate that the recycling decision is sensitive to perceived time costs, implying that own-cost elasticity estimates for recyclables generation must incorporate the participation decision. While recyclers were asked how long it took to recycle a unit of material (the τ_j for a tin can, a newspaper, etc.), non-recyclers were not asked this question. To estimate time costs for non-recyclers, we follow Lee (1976) by estimating maximum likelihood Heckman selection models predicting the time costs for paper, glass and tin (Table 5). These models are "unit time cost functions" and are influenced by variables capturing revealed household production behavior (INTERVAL and BROUGHT) and other factors. INTERVAL measures the number of days spent accumulating recyclables before delivery to the convenience center, while BROUGHT is a dummy variable indicating if the respondents had recyclables in hand at the time of the interview. Time costs were positively related to INTERVAL and negatively related to BROUGHT. Home owners had lower stated time costs, while time costs were positively related to TIME, the four-point variable capturing the general notion of time costs. Predicted time costs were estimated by substituting observed values for all independent variables for recyclers. The same was done for non-recyclers except for INTERVAL, which was not observed. In this case, the mean INTERVAL for recyclers who did not bring recyclables to the center was substituted.

Time costs were estimated for paper, glass, and tin. Aluminum was eliminated from the analysis because so many respondents brought aluminum to a buyback center. The maximum likelihood time cost model for plastic did not converge, and the corresponding OLS Heckman model for plastic resulted in numerous negative predicted time costs. Given these poor theoretical and econometric results, plastic could not be included in the generation models.[13] Implications of the exclusion of aluminum and plastic are discussed below. Paper, glass, and tin, however, account for nearly 92% of the recyclable materials waste stream by weight. Since SWMPs are designed in terms of waste stream weight, the generation models will provide elasticity estimates sufficient for policy analysis.[14]

Predicted per unit time requirements for each household for each material were converted to a per pound requirement.[15] Individual material requirements were then averaged to obtain a time cost index for an "average" pound of recyclable material. This was then scaled by the wage rate (opportunity cost) to obtain the recycling cost per pound of material. The hourly wage rate was obtained by dividing predicted annual household income by 2000 hours.[16] Conversion factors and the range of predicted unit time costs are reported by material in Table 6.

Second-Stage Estimation: Household Generation

The generation models reported in Table 7 are estimated using maximum-likelihood. The dependent variable at the selection stage is a 0/1 variable indicating whether the household recycles any of the three materials. Dependent variables at the generation stage were measured at the household level (pounds per household per month) and at the per capita level (pounds per person per month). This second measure was obtained by dividing monthly household generation by the number of household members (HHMEMBER). Overall, the generation models performed quite well. The selection-stage specification was based on the models presented in Table 5, with variable coefficients retaining the same sign and statistical significance as in the individual material models. The inverse mill's ratio calculated from the selection model is used in the generation model, where it is highly significant (σ). This suggests that models of recyclables generation should not be estimated without adjusting for selection effects.

The recycling cost (RECCOST) and income terms in the generation models have the expected signs and are statistically significant ($p<0.05$ for all models).[17] In the household generation models, the weight of recyclables was insensitive to the number of people in the household (HHMEMBER). This is probably due to the large portion of newsprint in the weight of total recyclables. Since most households receive only one newspaper, the lack of statistical significance is understandable. The chi-square statistics reported in Table 7 test the restriction that the cost and income slopes equal zero. Failure to reject these restrictions would suggest that generation of recyclable material is independent of recycling time cost and income. The restrictions are rejected at p-values of less than 0.01. The models thus provide relatively strong evidence that an own-cost effect is present in the generation of recyclable material.

Elasticities cannot be calculated from the models using TIME at the selection stage because we cannot measure the impact of changing cost on non-recyclers. Instead, the linear generation models were re-estimated using RECCOST at the selection stage (specifications 3 and 4). Expected monthly generation and own-price arc elasticity estimates were then calculated (Table 8). The linear household model using RECCOST at the selection stage underestimated the mean recyclables weight (including all non-recyclers) by 7.1%, while the per capita model underestimated actual mean generation by 3.8%. The arc-elasticity was estimated by first calculating the expected generation for each household at the observed values for the independent variables, then re-calculating the expected generation at a 10% lower cost (i.e., 10% lower time requirement). Household recyclables generation proved to be elastic, with an estimated own-price arc-elasticity of -2.00. The corresponding measure for per capita generation was -1.69. If a 10% drop in time costs was available to all households, estimated mean monthly generation of recyclable material would increase by nearly 3 pounds per household, from 13.88 pounds per household to 16.72 pounds per household. A similar drop in cost would increase estimated per capita generation by 1 pound per person, from 5.94 pounds per month to 6.96 pounds per month.[18]

It is interesting to note the effect of the selection/tobit approach used to model the generation of recyclables. Households already recycling prove to be the least price responsive of the three groups of households. This is intuitively appealing because these are experienced recyclers who are already relatively efficient in the production of recyclables. Households that do not currently recycle are the most elastic, because these households move from a corner solution (no recycling) to an interior solution (recycling) as the price falls.

V. Concluding Comments

This paper has developed a household production model describing recycling behavior. In contrast with past research, an own-cost elasticity measure is reported. This is important because rural regions of the country face increasing disposal costs, but may not find it feasible to use unit pricing for garbage as an incentive to decrease the amount of material entering landfills or incinerators. The empirical models describing participation in recycling and generation of recyclable material were consistent with the theoretical model and provided a number of insights into the design of recycling programs. Taken together, the participation and generation models present mixed results for rural recycling programs. At first glance, the participation models might lead one to conclude that recycling programs are not effective in inducing households to recycle. The availability of information about recycling and making households aware that revenues from the sale of recyclable material are donated to local community groups in the vicinity of the dropoff site do not appear to influence participation. Educational programs aimed at school-aged children also do not appear to have impacted household participation rates. Our models do not, however, provide a clear assessment of the impact of these program efforts. A more powerful test of program effectiveness would employ variables specifically designed to capture program effects.

The generation models, however, indicate that the weight of recyclables is elastic with respect to time costs. Thus, programs aimed at reducing household perceptions of the time costs of recycling or helping households become more efficient recyclers may well pay dividends in the form of increased participation and generation. These results are based on only three of the five materials accepted at the recycling centers. The two materials eliminated from the analysis - plastic and aluminum - are the lightest materials (relative to volume) and consequently would represent the materials with the highest per pound recycling costs. Using the four material (OLS time requirement) model as a guide, including these materials would adjust the elasticity estimates for an average pound of recyclables slightly upwards (less elastic).

The estimated average household generation of all recyclables, including non-recyclers, is about 15 pounds per month. This is obtained in a rural area where the recycling system is simply a series of dropoff sites rather than more costly curbside collection. Thus, it appears that dropoff programs can divert a substantial portion of the waste stream, particularly if some of the savings from a lower-cost dropoff program could be effectively used to lower the perceived time costs of potential participants.

ENDNOTES

1. If the amount of material recycled also yields utility, the utility function may be specified as U(Z, G, L, R), where R is the amount recycled and $U_R>0$. The first order conditions imply that even more waste will be recycled. Unfortunately, they also imply that products are purchased for the opportunity to recycle (Morris and Holthausen).

2. The derivatives with respect to H and the LaGrangian multipliers are suppressed.

3. At an interior solution in the labor market, μ/λ equals the wage rate (Shaw, 1992).

4. For ease of exposition, an implicit assumption is that the consumption commodity Z(x) and trash production T(x) require zero time in production. The model is easily extended to include time requirements in these technologies.

5. If consumers are willing to pay a premium to obtain packaging with particular characteristics, trash available for recycling is not costless. Assuming the price of x_j is evenly distributed across product and packaging, the cost function would then be,

$$c_j = (\omega/\tau_j) + (\theta_j p_j)/\gamma_j$$

The recycling technology also assumes an absence of other fixed or variable costs since recyclables require little capital investment and may be collected in paper or plastic bags in which groceries were once carried. Time costs remain the bulk of recycling costs.

6. All but one person interviewed combined recyclables delivery with other activities (e.g., garbage delivery or a shopping trip). To allocate the travel costs between recycling and other activities would prove very difficult. Because travel costs are very small relative to time costs this is not a major empirical problem.

7. Respondents were interviewed at three of the seven rural convenience centers. Interviews were conducted at convenience centers rather than at individual homes because the goal was to focus on those households with no trash hauling services. While this kept survey costs relatively low, the methodology introduces endogenous stratification into the sample, i.e., interviews are more likely to occur with those who make frequent visits to convenience centers rather than those who visit infrequently. To adjust for this

problem, it would be necessary to know the relationship between the sampling method and the probabilities of visitation for each individual (Shaw, 1988). This was not possible given the data available.

8. Four observations (less than 2% of the sample) were dropped because of extremely high generation rates for more than one material.

9. Additionally, many recyclers brought aluminum to a local commercial buy-back center.

10. A referee has noted that the respondent is the individual hauling the recyclables, and may not be the household member responsible for recycling decisions. In this case, the respondent's characteristics act as proxy variables.

11. Because the continuous income variable is predicted with error, a generated regressors problem exists. The reported second stage standard errors have been adjusted accordingly (Murphy and Topel, 1985).

12. A truncated generation model for just those bringing a positive amount of material would be inappropriate since we do know something about non-recyclers. Employing a standard Heckman two-step procedure would require us to drop the large number of "fortuitous" zero-generation observations.

13. A number of alternatives to dropping plastic were considered and rejected. First, all negative price observations could be dropped, but a substantial portion of the sample would be lost. Replacing negative values with zeroes is clearly inappropriate since this sets cost equal to zero. Although the Heckman modeling strategy for the cost functions is the theoretically correct approach, a four material model was estimated using OLS cost functions, which eliminated the negative predicted price problem. Elasticity estimates were not substantially affected (note #18).

14. Weight is an imperfect measure for solid waste and recyclables. Volume may be a more relevant indicator for some policy purposes (e.g., saving landfill space), but legislation is most often written for aggregate weight.

15. Interviewers noted the modal container size for glass and tin. The weight of the modal size was used as an estimate of $\gamma_j \theta_j x_j$. The average daily weight of the local newspaper was used for paper.

16. This is a common approximation for the opportunity cost of time. Unfortunately, the opportunity cost of time departs from the wage rate under a wide variety of conditions: if the number of work hours is not

freely chosen, if non-wage income is large relative to wage income, if work directly yields utility, if the marginal wage is non-linear, and if taxes are non-zero [Shaw (1992); McConnell and Strand (1981); Bockstael, Strand and Hanemann (1987)]. The data do not permit detailed investigation of all possibilities.

17. The generated regressors problem is also present in these models, but is far more complicated because the cost variable is the product of two first stage estimates. While adjustments to the variance-covariance matrix in this complex case are beyond the scope of this paper, the results from the participation models may be informative. For example, the standard errors on the cost and income parameters in specification 1 of Table 7 would have to double before the parameters became insignificant. In contrast, the standard errors adjustments in specifications 1 and 2 of Table 4 ranged between a 0.2% decrease and a 30.7% increase.

18. A four material model (including plastic) was estimated using the OLS time cost functions. The own-price arc-elasticity measure for the household model is -1.69, while the per capita elasticity estimate is -1.46.

REFERENCES

Bockstael, N.E., I.E. Strand, and W.M. Hanemann. 1987. "Time and the Recreation Demand Model," *American J. Agricultural Economics*, 69(2), 293-302.

Catsiapis, G. and C. Robinson. 1982. "Sample Selection with Multiple Selection Rules." *J. Econometrics*, 18(3):351-368.

Darcey, S. 1991. "Landfill Crisis Prompts Action," *The Management of World Wastes*, 32(5), 28-55.

Duggal, V.G., C. Saltzman, and M.L. Williams. 1991. "Recycling: An Economic Analysis," *Eastern Economic J.* 17(3), 351-358.

Fullerton, D. and T.C. Kinnaman. 1993. "Household Demand for Garbage and Recycling Collection with the Start of a Price per Bag." Paper presented at the Annual Meeting of the Allied Social Sciences Association, Boston, MA, January 1994.

Greene, W. 1991. <u>LIMDEP Version 6: User's Manual and Reference Guide.</u> Bellport, N.Y.: Econometric Software, Inc.

Hanemann, W.M. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American J. Agricultural Economics*, 66(2), 332-341.

Hong, S., R.M. Adams and H.A. Love. 1993. "An Economic Analysis of Household Recycling of Solid Wastes." *J. Environmental Economics and Management*, 25(2):136-146.

Lee, L.F. 1976. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19(2), 415-433.

Lee, L.F. 1983. "Generalized Econometric Models with Selectivity." *Econometrica*, 51(2):507-512.

McConnell, K.E. and I. Strand. 1981. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sportfishing," *American J. Agricultural Economics*, 63(1), 153-156.

Morris, G.E. and D.M. Holthausen. 1994. "The Economics of Household Solid Waste Generation and Disposal." *J. Environmental Economics and Management*, 26(3), 215-234.

Murphy, K.M. and R.H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models," *J. Business and Economic Statistics*, 3(4), 370-379.

Pollack, R.A. and M.L. Wachter. 1975. "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," *J. of Political Economy*, 83(2), 255-277.

Pudney, S. 1989. <u>Modelling Individual Choice.</u> Cambridge, MA: Basil Blackwell.

Shaw, D. 1988. "On-Site Samples' Regression: Problems of Non-negative Integers, Truncation and Endogenous Stratification," *J. Econometrics*, 37(2), 211-223.

Shaw, W.D. 1992. "Searching for the Opportunity Cost of an Individual's Time," *Land Economics*, 68(February), 107-115.

Stewart, M.B. 1983. "On Least Squares Estimation When the Dependent Variable is Grouped," *Review of Economic Studies*, 50(4), 737-753.

Table 1: Variable Means

Variable	Mean or %	Std. Dev.	n
INCOME ($1000)	39.85	22.16	262
EDUCATION (Years)	12.45	3.35	284
GENDER (% Female)	27.1	0.45	284
OWN HOUSE (% owning)	85.56	0.35	284
HHMEMBER (# of household members)	3.08	1.39	284
AGE (Years)	44.38	14.89	283
FRIENDS: I have friends who recycle. (% yes)	74.32	0.44	257
DONATE: Before this interview, I knew that recycling revenue went to local groups. (% yes)	31.34	0.46	284
CHILDREN 6-14: My household has children aged 6-14. (% yes)	29.22	0.46	284
Other variables measured on 4-point scale: 1 = Strongly Agree 2= Agree 3 = Disagree 4 = Strongly Disagree			
TIME: It takes little time to recycle.	2.02	0.53	274
STORAGE: My house has adequate storage space for recyclables.	2.49	0.74	284
GENERATE: My household generates enough material to make recycling worthwhile.	2.14	0.69	281
INFORMATION: Information on recycling is easily available.	2.13	0.54	265

Table 2: Participation and Generation Statistics

	Paper	Plastic	Glass	Aluminum	Tin	All
% saying they recycled this material (n=280)	54.3	52.5	39.6	61.8	30.4	74.6
% actually bringing the material to the center	21.8	15.7	11.1	10.4	10.7	30.7
Mean Monthly Generation (lbs. per month, all observations)[a]						
Household	10.55	0.72	2.38	0.55	0.82	15.04
Per Capita[b]	4.42	0.27	1.04	0.20	0.32	6.26
Mean Monthly Generation (lbs. per month, only those who brought the material)[c]						
Household	46.15	4.59	21.54	5.28	7.65	48.79
Per Capita	19.41	1.73	9.38	1.90	3.01	20.30
n	61	44	31	29	30	86
Maximum Monthly Generation (lbs. per month)[d]						
Household	139.40	13.07	78.43	24.40	26.14	182.10
Per Capita	61.00	6.54	33.89	8.71	13.07	85.40

[a] Calculated using all 280 usable observations, including recyclers and nonrecyclers whose weights were zero.

[b] Household recyclables weight divided by number of household members (HHMEMBER).

[c] Calculated using only those observations bringing recyclable material to the convenience center.

[d] The maximum monthly generation observed in the sample of observations bringing the material.

Table 3: Predicted Income Model

Intercept	-18.46 (-1.21)
AGE	1.35* (2.67)
AGE Squared	-0.02* (-3.08)
OWN HOUSE (1 = Yes)	13.17* (3.81)
GENDER (1 = Female)	-4.60 (-1.68)
RACE (1 = Non-White)	-6.26 (-0.82)
EDUCATION (Years)	-1.10 (-0.60)
EDUCATION Squared	0.22* (2.67)
σ	19.27* (20.63)
χ^2	116.94*
n	274

Numbers in parentheses are ratios of a coefficient to its asymptotic standard error.
* Significant at α=0.05.

Table 4: Recycling Participation Models
Dependent Variable: 1 = Recycle the material, 0 = Don't recycle the material

Variable	#1	#2	#3
Intercept	0.92 (0.93)	1.00 (1.07)	0.90* (2.24)
Household Production Technology and Characteristics Variables			
FRIENDS	1.04* (3.99)	1.07* (4.12)	
STORAGE	-0.50* (-3.09)	-0.50* (-3.16)	
GENERATE	-0.40* (-2.25)	-0.35* (-2.05)	
TIME	-0.38 (-1.79)	-0.41* (-1.96)	
AGE	0.02* (2.19)	0.02* (2.51)	
EDUCATION	0.07 (1.58)	0.08 (1.67)	
INCOME	0.35×10^{-4} (0.01)	0.44×10^{-4} (0.06)	
County Recycling Program Variables			
CHILDREN 6-14	-0.18 (-0.70)		-0.34 (-1.65)
DONATE	0.22 (0.86)		0.38 (1.76)
INFORMATION	0.13 (0.62)		-0.10 (-0.59)
n	209	209	209
χ^2	64.04*	62.40*	6.03
% Correct	80.4	79.5	75.1

Numbers in parentheses are ratios of a coefficient to its asymptotic standard error. Models with INCOME generated at the first stage have corrected standard errors (Murphy and Topel, 1985).

* Significant at $\alpha = 0.05$

Table 5: Unit Cost Functions

	Paper	Glass	Tin
Selection Model Dependent Variable: 1 = Recycle the material, 0 = Don't recycle the material			
Intercept	2.27*	0.71	0.74
	(3.72)	(0.88)	(1.03)
FRIENDS	0.82*	1.37*	1.40*
	(3.65)	(4.16)	(3.49)
STORAGE	-0.36*	-0.31*	-0.36*
	(-2.97)	(-2.07)	(-2.34)
GENERATE	-0.57*	-0.54*	-0.64*
	(-3.60)	(-2.91)	(-2.92)
AGE	0.02*	0.03*	0.01
	(2.44)	(3.72)	(1.59)
TIME	-0.65*	-0.65*	-0.37
	(-3.66)	(-2.69)	(-1.42)
n	238	230	228
Cost Function Dependent Variable: Seconds to recycle one unit of material			
Intercept	32.03	44.05	61.60
	(1.15)	(1.11)	(0.78)
INTERVAL (Days)	0.42*	1.77*	0.15
	(3.40)	(3.58)	(0.12)
BROUGHT (1 = Yes)	-10.47	-19.07	-12.34
	(-0.87)	(-1.06)	(-0.29)
OWN HOUSE (1 = Yes)	-29.79*	-25.47	-13.89
	(-3.04)	(-1.33)	(-0.32)
TIME	12.36	1.48	2.92
	(0.79)	(0.08)	(0.06)
σ	47.98*	66.18*	72.04*
	(20.70)	(14.76)	(5.98)
ρ	-0.08	0.03	-0.20
	(-0.19)	(0.08)	(-0.28)
n	136	97	74
Predicted Mean Time, All (Seconds Per Unit)	37.10	50.78	49.44
Recyclers	31.44	46.67	40.65
Non-recyclers	44.63	53.78	53.67

Numbers in parentheses are ratios of a coefficient to its asymptotic standard error.
* Significant at $\alpha = 0.05$.

Table 6: Unit Conversion Factors and Recycling Costs

	Units/Lb. of Material $[1/(\gamma_j\theta_j x_j)]$	Range of Costs, $/Lb. $\$/Lb. = [\omega_i/\tau_{ij}]*[1/(\gamma_j\theta_j x_j)]$		
		Recyclers	Non-Recyclers	All
Paper	1.3 Newspapers/lb.	0.01 - 0.73 (n=128)	0.04 - 0.85 (n=94)	0.01 - 0.85 (n=222)
Glass	1.33 Jars/lb.	0.05 - 1.31 (n=95)	0.09 - 0.97 (n=126)	0.05 - 1.31 (n=221)
Tin	7.69 Cans/lb.	0.34 - 4.80 (n=73)	0.34 - 5.36 (n=146)	0.34 - 5.36 (n=219)
RECCOST		0.15 - 2.19 (n=133)	0.25 - 2.35 (n=75)	0.15 - 2.35 (n=208)

Table 7: Generation Models

Specification	#1 HH[a]	#2 PC[b]	#3 HH[a]	#4 PC[b]
Selection Model (α) n=208				
Dependent Variable: 1= Recycle any of 3 materials, 0 = Don't recycle these materials				
Intercept	2.94* (3.56)	2.84* (3.49)	1.61* (2.42)	1.60* (2.41)
FRIENDS	1.28* (5.04)	1.27* (4.86)	1.23* (4.83)	1.21* (4.69)
STORAGE	-0.60* (-4.40)	-0.58* (-4.25)	-0.59* (-4.36)	-0.58* (-4.28)
GENERATE	-0.71* (-4.20)	-0.72* (-4.22)	-0.72* (-4.43)	-0.73* (-4.47)
AGE	0.02* (3.01)	0.03* (3.06)	0.03* (3.17)	0.03* (3.19)
TIME	-0.72* (-3.53)	-0.69* (-3.41)		
RECCOST			-0.18 (-0.82)	-0.17 (-0.78)
Generation Model (β) n=133				
Dependent Variable: Weight of recyclables brought to dropoff center				
Intercept	-18.64 (-0.97)	-9.14 (-1.39)	-15.77 (-0.82)	-7.86 (-1.17)
RECCOST (β_1)	-102.11* (-4.16)	-37.02* (-2.92)	-98.55* (-3.94)	-35.72* (-2.77)
INCOME (β_2)	2.46* (4.49)	0.93* (3.57)	2.38* (4.24)	0.90* (3.37)
HHMEMBER	-1.02 (-0.21)		-0.92 (-0.19)	
σ	59.30* (8.56)	26.66* (9.88)	59.32* (8.50)	26.77* (9.71)
ρ	0.07 (0.27)	-0.12 (-0.47)	-0.06 (-0.25)	-0.22 (-0.83)
$\chi^2(\beta_1=\beta_2=0)$	18.74*	13.58*	17.54*	12.78*

[a]Household Model [b]Per Capita Model
Numbers in parentheses are ratios of a coefficient to its asymptotic standard error.
* Significant at α=0.05.

Table 8: Mean Expected Generation and Arc-Elasticity Estimates

	Household Model	Per Capita Model
Predicted Generation (lb.) (from model)	13.88	5.94
Actual Generation (lb.)	14.86	6.18
Elasticity		
Overall (n=208)	-2.00	-1.69
Recyclers bringing material (n=68)	-1.82	-1.53
Recyclers bringing no material (n=86)	-2.01	-1.69
Non-Recyclers (n=54)	-2.21	-1.91

USING BEST MANAGEMENT PRACTICES TO CONTROL AGRICULTURAL WASTE RUNOFF: PUBLIC BENEFITS AND POLICY IMPLICATIONS

John C. Bergstrom, Hugh F. MacDonald, and Jack E. Houston*

*Bergstrom and Houston are Associate Professors, Department of Agricultural and Applied Economics, The University of Georgia, and MacDonald is a Research Associate, Department of Resource Economics and Policy, University of Maine.

USING BEST MANAGEMENT PRACTICES TO CONTROL AGRICULTURAL WASTE RUNOFF: PUBLIC BENEFITS AND POLICY IMPLICATIONS

Abstract

Nonpoint source water pollution from agricultural operations is a major concern in the agricultural and environmental policy arenas. The study reported in this paper estimated the off-site or off-farm benefits to the general public of using "best management practices" (BMPs) to reduce livestock waste run-off from dairy farms. The study area was the Little River/Rooty Creek watershed in the State of Georgia. Study results indicate that the off-site benefits of BMPs to Georgia citizens has a lower bound of about $220 million annually. This value estimate can be used by federal and state agricultural and resource management agencies in benefit-cost analyses of policies and programs designed to encourage the implementation of BMPs to control water pollution problems.

USING BEST MANAGEMENT PRACTICES TO CONTROL AGRICULTURAL WASTE RUNOFF: PUBLIC BENEFITS AND POLICY IMPLICATIONS

INTRODUCTION

One of the largest concentrations of dairy farms in the state of Georgia is found in the Little River/Rooty Creek watershed. This watershed, with a drainage area of approximately 218,000 acres is located in Jasper, Newton, Walton, Morgan, and Putnam counties. Waste (e.g., manure) runoff from dairy farms into water bodies in the watershed is a potential source of water quality problems. A consortium of federal and state agricultural and natural resource management agencies lead by the U.S. Agricultural Stabilization and Conservation Service and the U.S. Soil Conservation Service embarked on a joint project to study the feasibility of using "Best Management Practices" (BMPs) to control waste runoff from agricultural operations. The main objective of this portion of the joint project was to quantify the off-site public benefits provided by BMPs in the Little River-Rooty Creek watershed. Such information is important for evaluating the desirability of proposed public projects. Monetary estimates of the benefits provided by BMPs can be useful for justifying expenditures on public water quality improvement projects using BMPs, and to gauge the level of public support for such projects.

METHODS

Identification of Benefits

The on-site benefits of BMPs refer to the benefits of BMPs which accrue to the farmers who live and work on the farms where a BMP is located. For example, the use of a BMP on a particular dairy farms may help protect the quality of that farm's drinking water supply. Off-site or off-farm benefit of BMPs accrue to people who live, work, and recreate away from the farm or farms where BMPs are located. In the Little River-Rooty

Creekwatershed, the primary off-site benefit of BMPs designed to reduce dairy waste run-off is improved water quality in Lake Sinclair. Because of its location at the base of the watershed, much of the surface water in the watershed eventually drains into Lake Sinclair. Lake Sinclair provides recreational opportunities to local residents and others, and is used as a water supply for parts of Hancock county. Hence, the emphasis of this study was on the economic benefits of maintaining a specified level of water quality in Lake Sinclair.

Benefit Estimation

The off-site benefits provided by BMPs are in the nature of nonmarketed goods because these benefits are not traded as commodities in regular economic markets. Consequently, explicit (or, observed) prices for the off-site services provided by BMPs do not exist and money metric estimates of these benefits must be obtained by other means. Nonmarket valuation has been developed as an alternative approach for estimating the economic value of such benefits. Nonmarket valuation techniques have become increasingly popular for evaluating environmental policy and management alternatives, and have been accepted by the U.S. Water Resources Council for the assessment of water policies and projects. In this case study, a form of contingent valuation (CV) was used (Mitchell and Carson, 1989). CV involves the use of survey techniques to elicit values directly from individuals affected by the nonmarketed good in question. The survey also collects various household socioeconomic and preference data.

The CV questionnaire elicited individuals' annual willingness to pay (WTP) to support the use of BMPs for protecting water quality in Lake Sinclair. Included in the questionnaire was a description of the pollution problem in the Little River-Rooty Creek watershed and the proposed corrective action - establishment of selected BMPs (e.g., no till systems, animal waste stack houses, erosion or water control structures). In a referendum-type framework, respondents were asked if they would vote to support alternative water pollution control programs given that it would decrease their annual income by a stated amount of money. This type of valuation question is termed the dichotomous choice (DC) or closed-ended approach. An advantage of the DC approach for valuing nonmarketed goods is that it approximates the situation faced by consumers in regular markets. That is, when considering the purchase of a private good, consumers base their decision on the offer price of the good. The purchase is a "yes or no", or "take it or leave it", decision. The disadvantage of the closed ended approach is that less information is obtained from each individual.

The objective of CV is to determine an individual's *maximum* WTP for the nonmarketed commodity. Therefore, this value must be inferred from only the "yes" or "no" responses to the closed-ended questions. As such, a more complicated statistical analysis is required than would be the case if maximum values were elicited directly in an open-ended (fill in the blank) valuation framework. However, given the nature of the commodities being valued (water quality) it was felt that respondents could answer more meaningfully to a "yes" or "no" situation than having to determine their exact WTP for a good with which they had little experience in valuing.[1]

Respondents were informed that implementation of BMPs would ensure water quality in Lake Sinclair suitable for fishing and swimming. This type of water quality index is an adaptation of that used in a national study of water quality benefits (Mitchell and Carson, 1984). The responses to the valuation questions provide information on individual WTP for water pollution control using BMPs. Respondents were informed that technical and financial

[1]Copies of the survey questionnaire and technical details on the dichotomous choice CV approach used in this study are available from the authors upon request.

assistance would be provided to farms in the watershed to aid in the adoption of selected BMPs.

Survey Procedures

The questionnaire was pretested on 100 randomly selected households in the Lake Sinclair area in November, 1992. The pretest revealed the need for minor clarifications of certain questions and provided information on the range of household WTP for alternative pollution control programs. The formal survey was conducted in January and February, 1993. Based on the WTP data obtained from the pretest and a review of similar studies, the offer amounts for the WTP questions were $5, $10, $50, $75, $150, $300, $500, $750, $1000, and $1500. Each questionnaire was randomly assigned one of these values.

The questionnaire was mailed to 1900 randomly selected households in Atlanta and the Lake Sinclair area. The sampling plan was based on a license plate survey of lake users at various sites around Lake Sinclair, which identified a market area for the lake. Questionnaires were sent to ten counties around Lake Sinclair and nine in the Atlanta area. The counties chosen around Lake Sinclair were Baldwin, Bibb, Greene, Hancock, Jasper, Jones, Morgan, Putnam, Taliaferro, and Wilkinson. In the Atlanta area, the counties of Clayton, Cobb, Dekalb, Douglas, Fayette, Fulton, Gwinnett, Henry, and Rockdale were targeted. Mailouts were weighted by populations of selected counties for a given mailout region. For example, of the 1275 questionnaires sent to the Lake Sinclair region, 140 were sent to Baldwin County, given that Baldwin constitutes approximately 11% of the total population of the selected, ten-county region around Lake Sinclair. The survey was conducted in accordance with procedures advocated by Dillman (1978). Included with the questionnaire in the initial mailout, a cover letter explained the purpose of the survey and stressed the importance of an individual's response. A map showing the location of Lake Sinclair and the portion of the lake affected by agricultural nonpoint source pollution also accompanied each questionnaire. One week after the initial mailout, a reminder post card was sent to those who had not returned a completed questionnaire. A second and third mailout of questionnaires with maps and cover letters were sent two weeks and four weeks after the initial mailout to those who had not returned a completed questionnaire.

RESULTS

Response Rate

Of the 1900 questionnaires mailed, 325 were returned as undeliverable, leaving an adjusted sample size of 1575 (Table 1). Of 690 questionnaires returned, 18 questionnaires were returned blank, leaving an adjusted response rate of 42.7%. On a regional basis, the adjusted response rates for the Atlanta and Lake Sinclair areas were almost identical at 43.0% and 42.5%, respectively. These figures are within the range of response rates experienced in other CVM studies (Mitchell and Carson, 1989, p. 281).

Household Characteristics

Descriptive statistics of the households surveyed are presented in Table 2. Such information is important from an economic perspective, since household WTP and demand for public goods varies with individual household characteristics. The average gross annual income of the 672 households that completed the questionnaires was $50,217. At first glance, this number appears to be unexpectedly large. However, one would expect average annual income to be higher in the Atlanta area in relation to the average Georgia household. The results of this survey are indicative of this, as the average income of responding Atlanta area households was $60,081, while those in the counties around Lake Sinclair reported an average of $44,701. Given that approximately 35% of the data is based on Atlanta households, this estimate is reasonable. Education can be an important factor influencing an individual's WTP, since some of the benefits of enhanced water quality and wildlife habitat require a certain amount of abstraction and scientific understanding. Thirty-nine percent of respondents have university degrees. Respondents appeared to be familiar with Lake Sinclair; approximately 68% indicated that they had visited the lake on at least one occasion, and 20 respondents actually live on the lake.

Mean WTP for water pollution control was estimated at $315 per household per year. The 95% confidence interval for mean WTP is $223 (lower bound) to $413 (upper bound). The estimate of mean WTP represents the maximum annual WTP per household for the implementation of BMPs in the Little River-Rooty Creek watershed which would ensure water quality in Lake Sinclair suitable for fishing and swimming.

DISCUSSION AND IMPLICATIONS

The environmental impacts of agricultural operations is receiving considerable attention in the public policy arena. Interrelationships between agriculture and the environment was a major topic during the 1990 U.S. Farm Bill debate, and provisions for protecting environmental quality promises to be an area of even more interest and attention with respect to the upcoming 1995 U.S. Farm Bill debate. In addition, as the U.S. Clean Water Act is amended in the future, more provisions for controlling water pollution from agricultural operations are likely to be added. More and more state governments are also implementing or are considering implementing agricultural pollution prevention programs.

In the Little River/Rooty Creek watershed, waste runoff from the large number of dairy farms in the watershed is a water quality concern which has drawn the attention of federal and state agricultural and environmental agencies. These agencies are working together to examine the feasibility of using "Best Management Practices" to control waste runoff from dairy farms in the watershed. This program involves providing farmers with technical assistance and cost-sharing for implementing BMPs on individual farms.

What is the justification for using federal and state taxpayers' dollars to subsidize BMPs on private farms? It would indeed be difficult to justify such expenditures if the only beneficiaries are the farmers themselves. As argued in this paper, however, BMPs may provide broader off-site or off-farm benefits to the public in the form of improved downstream water quality. The strength of this argument in favor of public subsidies for BMPs depends on whether or not the public benefits provided by BMPs exceed the public costs.

At the lower bound of the 95% confidence interval, the economic value of the off-site public benefits of BMPs in the Little River/Rooty Creek watershed was estimated to be approximately $223 per household per year. This mean value estimate represents households in 10 counties around Lake Sinclair and 9 counties in the Atlanta metropolitan area. The total number of households in these 19 counties is approximately one million (Bachtel). The aggregate economic value of the off-site public benefits of BMPs in the Little River/Rooty Creek watershed is therefore conservatively estimated at about $223 million per year. This aggregate value estimate is on the conservative side for two reasons. First, the lower bound of the 95% confidence interval for mean WTP is used in the aggregation. Second, it is assumed that households in all other counties in the state of Georgia which were not sampled place a zero economic value on the off-site water quality benefits of BMPs in the Little River/Rooty Creek watershed.

The aggregate benefit estimate of $223 million per year suggests that the off-site benefits to Georgia citizens of using best management practices to control agricultural pollution runoff are substantial. This value estimate can be applied by federal and state agricultural and resource management agencies in benefit cost analyses of nonpoint pollution control policies and programs. The costs of best management practices include private construction and operation costs plus any government subsidies or incentives paid to farmers to implement best management practices. A useful extension of this study would be to conduct a "full-blown" benefit-cost analysis of best management practices which considers all relevant private and public benefits and costs. The results of this study would provide an important input into such an extended benefit-cost analysis.

REFERENCES

Bachtel, D.C. (Editor). The Georgia County Guide: Thirteenth Edition. The University of Georgia Cooperative Extension Service, College of Agricultural and Environmental Sciences. Athens, GA. 1994.

Dillman, D.A. 1978. **Mail and Telephone Surveys-The Total design Method**. New York: Academic Press.

Mitchell, R.C., and R.T. Carson. 1984. National fresh water quality benefits: Findings from a new national contingent valuation study. Draft Discussion Paper Presented at the Summer Meeting of the Association of Environmental and Resource Economists, Cornell University.

_____.1989. **Using Surveys to Value Public Goods: The Contingent Valuation Method.** Resources for the Future, Washington, D.C.

Table 1. Summary of Survey Response Rate, Study of Off-Site Benefits of Best Management Practices for Controlling Agricultural Waste Run-off.

	Atlanta Area	Lake Sinclair Area	Total	Ratio (%)
Total Mailed	625	1275	1900	
Undeliverable	95	230	325	17.1
Adjusted Sample Size	530	1045	1575	
Total Returns	238	452	690	43.8
Blank Questionnaires	10	8	18	
Adjusted Returns	228*	444*	672	42.7

* Response rates for the Atlanta and Lake Sinclair areas are therefore 43.0% and 42.5%, respectively.

Table 2. Household Characteristics of Survey Respondents, Study of Off-Site Benefits of Best Management Practices for Controlling Agricultural Waste Run-Off

Characteristic	Mean (Percentage of Respondents)
Income	$50,217
University Degree	39.1%
Age	49.6%
Female Respondents	32.2%
People per Household	2.6%
Visited Lake Sinclair*	67.8%
Member of Environmental Org.	17.7%

* 20 Respondents live on Lake Sinclair

Valuing Rural Environmental Amenities: Modeling Considerations, and Preliminary Results from the NSRE.

Daniel Hellerstein

Peter Feather **

Presented at the March 1995 W-133 Meetings
Asilomar, California.

Abstract

The valuation of rural environmental amenities poses some unique problems. In particular, the highly dispersed nature of rural resources, with the affected population living within a variegated landscape, makes it difficult to apply standard, site-based, measurement techniques. In this paper, we review some of these issues, and suggest means of surmounting them.

** Daniel Hellerstein and Peter Feather are natural resource economists in the Natural Resources and Environment Division, Economic Research Service, United States Department of Agriculture. The view expressed in this article do not necessarily represent the policies or views of the U.S. Department of Agriculture.

Valuing Rural Environmental Amenities: Modeling Considerations, and Preliminary Results from the NSRE.

1. Introduction

Somewhere between the cities where most people live, and the dedicated wildlands that constitute primordial nature, lies the rural environment that comprises the bulk of America These landscapes are characterized by a broad mixture of land-uses, with some form of agriculture predominating. Commodity production (crops and livestock), while the primary set of economic outputs, is by no means the sole good flowing from these lands.

In particular, opportunities for outdoor recreation, while lacking the solitude and spectacle of the backcountry, are manyfold (USDA RCA 1989). Wildlife, both game and non-game, can be present in significant numbers. Lakes, streams, and wetlands can be important local resources. Open space, with some mixture of planted and naturally occurring flora, can satisfy aesthetic and other tastes .

Furthermore, the hand of man can have major impacts on the quality of these resources. Of crucial importance is how society manages it's agricultural production. Whether it be the control of soil erosion, the application rates of potentially damaging chemicals, or the provision of undisturbed habitat; farm-level choices can have profound impacts on the ability of rural landscapes to provide an appealing natural environment.

2. Problems in the Valuation of Rural Environmental Quality

Upon recognizing the compelling logic of the above reasoning, the applied cost-benefit analyst is faced with an aggravating problem: the dispersed nature of the resources is not easily accommodated by standard measurement techniques. Rural environmental resources, almost by definition, are likely to have many spatial substitutes. The use of site-based travel cost models is thereby complicated, especially since the locales at which the resources are obtained (that is, the sites visited) are often difficult to identify. Furthermore, given the multiplicity of sites, and the informal means of access control often used, user based surveys would be very expensive to implement on a sufficiently broad scale.

Consider the problem of valuing the wildlife benefits of the Conservation Reserve Program (CRP). The CRP is a 36 million acre land set-aside, wherein farmers are paid a yearly fee to let the land stay fallow over an extended period (CRP contracts are for 10 to 15 years) (Osborn et al). Originally intended to protect high-erodible lands from tillage (with associated water and air quality benefits), and to reduce commodity surpluses, the CRP has also stimulated a boom in wildlife populations, especially birds (such as pheasants and migratory waterfowl) (Johnson et. al., Allen). Granting that more wildlife and clean water are probably good things, how then to measure the resulting benefits (or, in partisan speak: can we justify the CRP given it's $1.8 billion cost to the taxpayer)? Or, of more poignant concern, what guidance can be given as to the optimal size and geographic distribution of the CRP?

One possible approach would involve a large contingent valuation study, focused on potential users of the rural environment. Surveyee's would be asked to provide a value for alternative landscapes, with and without the beneficial influence of a CRP-like set-aside; that is, complete with summary measures of the qualities of resources likely to matter (i.e.; water clarity across all sites, overall wildlife populations, etc.) under different land use regimes. Although having the advantages of succinctness, such an approach is complete with the all-too-familiar controversies surrounding CV (Portney, Hanemann, and Diamond and Hausman). In this paper, we buck a trend and focus on revealed preference techniques, leaving discussion of CV techniques for a later date.

Another approach, squarely within revealed preference, would use hedonic housing price/wage measures (Blomquist et al) in conjunction with summary measures of rural environmental quality. If the attractiveness of rural landscapes, especially those features likely to be influenced by land retirement, is best thought of as contributing to one's ambient environmental quality, such an approach is probably optimal.

However, if rural environments serve as a destination for citified inhabitants (perhaps small cities), hedonic measures will lack power. Even in cases where the "market area" of specific chunks of land is localized (that is, not serving as a far away destination), the data resources are problematic. In particular, rural property values, although oft surveyed (USDA 1994, and Palmquist), are contaminated by such extraneous factors as differential land quality, and the presence of farm buildings (and other infrastructure). Again, for this moment we leave such measurements to others.

We are left with a consideration of how to apply site based methods, or more precisely, travel cost techniques. In the tool box of travel cost techniques, three main categories are relevant:

a) Hedonic travel cost models (Brown and Mendelsohn). For rural environmental attributes, this may be relatively easy to conceptualize; the broad dispersion of resources (with some regions more richly endowed then others) makes it likely that individuals are faced with different gradients for different attributes (say, secci disk of lakes).

b) Multiple site travel cost (Burt and Brewer). For this, we would need to classify sites into a discrete and finite set of classes; say, "unpolluted, mid-size lakes", or "slightly polluted large rivers". A nice feature of multiple site models is that they allow many visits to many sites.

c) Discrete choice/continuous (Bockstael et al). Given the broad range of attributes that may be relevant, and the fact that both choice of trip and number of trips are likely to be influenced by attributes, the discrete/continuous method is appealing.

In addition to the inherent strengths and weaknesses of these different classes of models, it is useful to review the "sampling" problems relevant to the valuation of rural environmental resources. These include:

1) How to identify sites. There may be vast number of small lakes and streams, much less fallowed fields, whose attractiveness can vary significantly. This is magnified when site quality measures are required, so that mere (albeit sometimes hard to obtain) knowledge of location is insufficient.

2) How to account for substitutes. Even if one knew what sites were visited by a sample, and had good measures of the qualities of these sites, the accounting for substitutes can be an onerous task; if for no other reason then the proper set of substitutes will include sites which may not appear as a destination anywhere in one's sample.

3) How to gather a sample. Surveying on-site a representative sample offers some hope, but still leaves the problem of estimating total demand and can be very expensive if there are many sites, each lightly visited. General population surveys give a measure of total demand, but run a foul of the identification (in the geographic sense) problem.

3. A Case Study: Valuation of the CRP using Available Data

Let us now return to the CRP valuation question. The most poignant feature of the CRP is it's vast dispersion, making impractical any attempt at an on-site survey (say, of wildlife-related activities). Furthermore, off-site consequences (say, erosion impacts on watersheds) are likely to be difficult to pinpoint, due to the lack of good broad-scale geophysical models.

To address these and other difficulties, we chose to aggregate sites. As a compromise between practicality and descriptive accuracy, we chose to use NRI Polygons (Kellogg et all). NRI Polygons are defined using county boundaries, Major Land Resource Area (MLRA) boundaries (USDA/SCS, 1981), and 8 digit hydrological code boundaries. The intersection of these 3 boundary sets yields about 14,000 discrete areas; so that each area is within the same county, MLRA, hydrological unit.

The main reason for this particular level of aggregation (say, as opposed to a town level aggregation) is the ability to match points in the National Resources Inventory (NRI) survey (USDA/SCS, 1984) to NRI polygons. The NRI consists of about 800,000 point samples of land use, land cover, erosion, and soil type. It, and related databases, have been used to create a series of environmental indices (Heimlich). Examples of these indices include delivered sediment to area streams, pesticide and nitrate leaching potential, and measures of habitat diversity.

By aggregating to NRI polygons, we avoid the necessity of identifying the exact location of a visited site, and the collection of qualitative information at the site level. Of course, we introduce an aggregation bias, both in terms of location (the NRI Polygon centroid being used instead of the actual site's location), and quality (depending on the homogeneity of the NRI polygon).

This still leaves the problem of acquiring demand information. We use two surveys to supply this data: the National Survey of Recreation and the Environment (NSRE) and the

National Survey of Fishing and Hunting, and Wildlife Associated Recreation (FHWAR).

The FHWAR is a well-known contact/recontact survey of about 100,000 people conducted by the US Fish and Wildlife Service. It has the advantage of large sample sizes, but focuses on broad choices and expenditures, without details on choices (although some contingent valuation and last trips questions are included to address some of these needs).

The NSRE is a multi-agency effort currently underway (due to be completed this spring) that surveys 20,000 people regarding their recreational choices. ERS has a small component of this (1500 observations), in which we focus on the water based recreational choices of individuals in 4 regions of the country: the lower Susquehanna watershed in Pennsylvania, the White River watershed in Indiana, Central Nebraska, and the Columbia River watershed near Spokane.

In particular, we ask the name and mileage/ordinal direction (N, NE, etc.) of up to 9 different waterbodies visited within the last year for any purpose. With this mileage and direction information, and using waterbody name (when the name found on a 1 to 500,000 USGS map), we assigned each visit to an NRI Polygon.

With this visitation information, we are currently estimating a discrete choice model, where the choices are all polygons within 120 miles of the home (zip-code) of each respondent, and the choice is the polygon actually visited. Note that each surveyee can contribute up to nine observations (for up to nine visited waterbodies) to the regression, and that each observation is weighted by the number of times the waterbody was visited (by the surveyee).

A second stage model of total trips taken will then be estimated, using either an inclusive value or a "probable" price model (Feather et al). From either model, consumer surplus numbers can be estimated; using policy scenarios that incorporate varying levels of CRP enrollment. Furthermore, these results can be extended nationwide, using zip code level socioeconomic variables, and the attributes of the NRI Polygons surrounding these zip codes.

Unfortunately, the NSRE dataset contains very little information on wildlife related activities (partially by design, since we did not want to replicate the FHWAR work). The FHWAR is an obvious choice for visitation information, but we can not obtain the same level of specificity on site visits. Instead, we will attempt a modified "gravity" model. Specifically, based on an individual's zip code, information on concentric bands around his origin will be constructed (say, average habitat indices within 25 miles, average habitat indices between 25 and 50 miles, and average habitat indices between 50 and 100 miles). These summary statistics (on the surveyee's local environment) will be used to estimate the number of trips taken by the surveyee. Variations in these summary statistics, as might occur under different levels of CRP, can then be used to estimate different levels of trip taking. This, when combined with average visitor day values, can be used to compute benefit numbers (see Hansen and Ribaudo for similar efforts).

4. Critique of CRP Modeling

This section is a plea for help. Clearly, our modeling strategy is a compromise. To show that we don't have our heads in the sand, let us review some of the more uniquely troublesome aspects of this strategy (ignoring generic problems that bedevil all travel cost demand analysis):

1) Aggregate Sites.

First and foremost, there is that troublesome notion of aggregate sites, implying that the quality of a site can be proxied by the qualities of the zone (i.e.; the NRI Polygon) within which it falls. As long as the zones are homogenous *across variables you care about* than this is may be tolerable. At the least, one might ask for tests of this supposition. Better yet, measures of within-zone heterogeneity could be used in expanded models (i.e.; Parsons and Needleman).

Another problem is deciding which zone to assign each "visited waterbody" to. Some waterbodies, especially rivers, cover several NRI polygons. We use self-reported distance and direction to select which polygon is correct; this is obviously subject to error (comparisons of locations generated by placenames versus thus from self-reported distance and direction are encouraging, but do suggest that moderate errors are quite likely). Thus, not only have we introduced an errors-in-variables due to aggregation, we also introduce one due to potential misassignation of zones.

A possible solution would be to obtain better place-name information.

Unfortunately, for the cases where no name is given (i.e.; "I went to A LAKE") or where the waterbody is large (i.e.; "The Columbia River"), this method won't be too helpful. Furthermore, one would still have to obtain site-quality information, which is not readily available (i.e.; river reach coverages are either absent, or present in too-fine a detail).

2) Environmental Indices.

Environmental indices are attempts to summarize attributes of a region. As mentioned above, for a point (a waterbody or a field) within this region, such as summary will be an error ridden measure. Furthermore, it is a once-removed (or more-then-once-removed) measure of the actual good being consumed. For example, if fish populations are important to fishing enthusiasts, or the look and feel of the water to boaters; it would be best to directly measure such variables (even as a zonal average). The use of these indices (i.e.; habitat suitability indices, delivered sediment indices, and nitrogen leaching indices) is justified to the extent that the factors comprising the indices also effect the goods of interest (eg. fish populations). Thus, we are essentially estimating a highly reduced form, with all the identification (in the simultaneity sense) problems therein (i.e.; equivalent changes in an index may have very different impacts on the goods of interest).

3) The benefits-transfer black box.

For the wildlife component of the model, the lack of locational information in the 1991 FHWAR suggests use of a gravity type of model predicting total trip takings. Valuation then requires a benefits transfer of per trip consumer surplus values. While a common strategy (i.e.; Ribaudo et al), it does leave much to be desired. In particular, the assumption that preexisting trips do not change in "average value", and that new (or reduced) trips have the same average value is highly dependent on model specification (Hellerstein).

5. Some Preliminary Results from the NSRE.

At this point we do not have extensive modeling results, so in this section we present some general statistics about recreational intensities, and some very early modeling results. Presented here are some early results for the freshwater recreation model using NSRE data.

5.1 Preliminary Random Utility Models

This section presents some preliminary random utility model (RUM) estimates for three of the four NSRE study areas[1]. On each outing involving two hours or less travel time, respondents reported where they went (destination name), type of destination (river, lake or wetland) and how far and in what direction they traveled. Using this information, and maps of each of the regions, each unique destination was assigned a latitude and longitude. The geographic locations of these destinations enabled them to be placed into an NRI polygons. Each study area has approximately 450 individual/destination pairings with an average trip frequency of approximately 6 trips per pairing. To compute travel cost, a distance matrix

[1]The three areas used are in Indiana, Nebraska and Pennsylvania. The fourth region (Washington) has poor NRI coverage and is not included in the analysis at this point.

was computed where travel distance is the straightline distance from each individual's zipcode centroid to the center of each NRI polygon. Estimating the models involves issues of aggregation and choice set definition that are discussed below.

5.2 Aggregation Issues

Although the use of NRI polygons as aggregate alternatives, rather than individual elemental alternatives, adds bias to the RUM estimation, some of the bias can be removed by incorporating a measure of size. Let the random utility of elemental alternative l to individual k $[U_{lk}]$ is written as:

$$(1) \qquad U_{lk} = V_{lk} + \epsilon_{lk}, \; l = 1,\ldots,N \qquad k = 1,\ldots,K$$

where V_{lk} is the deterministic portion of the utility function and ϵ_{lk} is an independently and identically distributed extreme value random variable with mode 0 and scale parameter μ. Now consider an aggregated model where disjoint sets $[L_i]$ of aggregate alternatives are formed from the set of élemental alternatives. The random utility of aggregate alternative i to individual k $[U_{ik}]$ is:

$$(2) \qquad U_{ik} = \mathrm{Max}(V_{lk} + \epsilon_{lk}) \quad \forall\, l \in L_i.$$

It has been shown [see Ben-Akiva and Lerman (1985)] that (2) can be decomposed into:

(3) $$U_{ik} = V^*_{ik} + (1/\mu)\ln(M_i) + (1/\mu)\ln(B_i) + \epsilon_{ik},$$

where V^*_{ik} is the average of the i-th aggregate alternative, M_i is the number of elemental alternatives in the i-th aggregate alternative, and B_i is a measure of the variability of the utilities of the elemental alternatives in the i-th aggregate alternative. Clearly, estimating (3) without the second and third terms (i.e. treating the aggregate alternatives as if they were elemental alternatives) introduces bias. Estimating an aggregated model using only V_{ik} results in bias unless either $1/\mu$ equals zero, or B_i and M_i are constant across aggregate alternatives. Neither of these conditions are likely. Because B_i is unknown, aggregate models will contain some bias even when $\ln(M_i)$ is included (see Feather, 1994).

To compound the aggregation problem, the number of potential elemental alternatives is unknown in each NRI polygon. As a proxy for the number of sites, kilometers of river length and kilometers of lake area were obtained from a 1:500,000 scale map for each NRI polygon. Lake area was used to roughly estimate lake shoreline length[2]. This results in consistent measures of size between lakes and rivers. The actual size measure then depends on which type of destination was visited. Specifically,

$$M_i = \begin{cases} \text{km river length if the respondent visited a river.} \\ \text{km lake shoreline length if the respondent visited a lake or wetland.} \end{cases}$$

[2]Assuming lake area represents one perfectly round lake, Area=PI*Radius². Solving for the radius allows for the shoreline, or circumference (2*PI*Radius) to be determined.

5.3 Choice Set Definition

Estimating RUMs requires knowledge of a universal choice set containing all relevant elemental or aggregated alternatives . Presently, there is no algorithm for determining the choice set. Researchers commonly gather as much information as possible about various destinations and let that set of alternatives represent the universal choice set. This is defensible only if each element in this choice set influences the decisions of each individual. Partitioning the universal choice set into individual specific choice sets (usually depending on distance from the individual's residence) is another commonly used tactic. Partitioning choice sets is especially appealing when survey questions involve trips within some allotted amount of travel time (e.g. trips to sites within x miles/travel time from your residence).

Since the NSRE survey questions deal with trips within two hours travel time from the respondent's residence, the partitioned choice set approach is used. The "full" choice set for each individual is all NRI polygons that are within 120 miles straightline distance from the centroid of the respondent's zipcode zone. Another approach would be to use a "reduced" choice set consisting of NRI polygons from the full choice set that are visited by one or more respondents is also used to estimate the RUMs. The reduced choice set could be justified on the grounds that it may more closely represents the choices that affect behavior than the full choice set. On the other hand, the full choice set contains information about numerous polygons that are never visited which could be potentially important. A further refinement of this model is likely to include a comparison of alternative choice sets and a nesting structure of lake/river choice followed by a site choice decision. Lake and river choice sets could be differentiated in terms of the existence of lakes and/or rivers in each NRI polygon.

5.4 Non-Nested RUM Results

The results of the non-nested RUM estimation appears below. The dependent variable is the number of trips taken to the destination. The parameter on the measure of size variable described above is constrained to one to be consistent with the assumption that the scale parameter of the error term in (3) is one. The independent variables are described below.

1. COST — Travel cost plus time cost in $1.00 units. Travel cost is the round trip straightline distance from the respondent's zipcode centroid to the NRI polygon centroid times $0.25. Time cost is the roundtrip distance divided by 50, then multiplied by one third on the respondent's annual personal income divided by 2000.

2. %CRP — Percentage of the total NRI polygon area that is occupied by CRP acreage.

3. %PROWN — Percentage of total NRI polygon area that is privately owned.

4. %FOREST Percentage of total NRI polygon area that is forested.

5. PEST Potential pesticide leaching index.

6. NITRO Potential pesticide leaching index.

7. EROS Estimated erosion (tons/acre/year)

Sine high levels of erosion are likely to be detrimental to water quality, the erosion parameter is assumed to be negative. Because both PEST and NITRO reflect areas where leaching of pesticides and nitrogen (resp.) could potentially be a problem, the anticipated signs on these parameters is negative. Since these indices only represent potential conditions, their correlation with actual conditions might not be accurate. The presence of forested land is assumed to be a environmental amenity, while the presence of private land is assumed to represent a lack of recreational opportunities. These variables are assumed to have a positive and negative parameter respectively. The parameter associated with CRP is assumed to be positive. Larger amounts of land in CRP, akin to lower levels of erosion, is assumed to increase water quality. Lastly, the travel cost parameter is assumed to be negative.

Results of the estimation appear in Table 1.

Table 1 -- RUM Estimates for Full Choice Set

Parameter	Indiana	Nebraska	Pennslyvania
COST	-0.1474 (-63.0)	-0.1771 (-66.4)	-0.1518 (-68.4)
%CRP	17.6989 (12.7)	9.0135 (10.7)	-35.7005 (-8.9)
%PROWN	-2.3054 (-16.6)	0.7936 (4.7)	0.8929 (6.7)
%FOREST	-1.0562 (-5.3)	-16.6718 (-15.1)	1.1146 (7.0)
PEST	-0.5443 (-3.1)	0.2846 (1.64)	-0.3142 (-9.9)
NITRO	-0.0481 (-7.1)	-0.0646 (-15.3)	-0.0231 (-3.2)
EROS	-0.3421 (-12.0)	-0.1810 (-11.1)	0.0264 (4.5)

6. Summary and Discussion

Non-market outputs of rural environments (i.e.; outdoor recreation) are potentially large. Government programs, such as the CRP, can have significant impacts on the size of these outputs. As such, estimation of the benefits of these non-market outputs, both for comparison against program costs and to compare alternative policies, is important.

The computation of these benefits is complicated by the dispersion of the rural environment and the rural population, a feature that renders the standard "origin to small set of sites" travel cost framework difficult to apply. In particular, the large set of relatively obscure sites introduces measurement problems, and also complicates efforts to control for substitutes. To deal with these difficulties, some form of aggregation is attractive.

We took advantage of the existence of a body of "environmental indicator" data, and aggregate at the NRI Polygon level. Using these sub-county zones, discrete choice models, with the "visited and other nearby NRI polygons" being the alternatives, are possible. Alternatively, gravity models, using zonal attributes to compute overall levels of participation are possible. We are currently investigating the effectiveness of these models, using NSRE and FHWAR data to account for the impacts of the CRP.

Some preliminary results, using NSRE data, suggests that this "zonal" approach may be practical. Needless to say, further work is required.

References:

Allen, Arthur. *"Regional and State Perspectives on Conservation Reserve Program Contributions to Wildlife Habitat"*. Report of the National Biological Survey, Midcontinent Ecological Science Center, Ft. Collins CO. (May 1994).

Ben-Akiva, M. and S.R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* (M.L. Manheim, editor). MIT Press Series in Transportation Studies, Cambridge, MA. (1985).

Blomquist, Glenn, Mark Berger and John Hoehn. "New Estimates of Quality of Life in Urban Areas". *American Economic Review*, 89 (March 1988), pp.89-105.

Bockstael, Nancy, W. Michael Hanemann and Catherine Kling. "Estimating the Value of Water Quality Improvements in a Recreational Demand Framework", *Water Resources Research,* 23 (1987), pp. 951-956.

Brown, Gardner and Robert Mendelsohn. "The Hedonic Travel Cost Method", *Review of Economic and Statistics*, 66 (1984), pp. 427-433

Burt, O. and D. Brewer. "Estimates of Net Social Benefits of Outdoor Recreation". *Econometrica*, 39 (1971), pp. 813-827.

Diamond, Peter and Jerry Hausman. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives*. vol. 8:4. Fall 1994.

Feather, Peter, Daniel Hellerstein and Theodore Tomasi. "A Discrete-Count Model of Recreational Demand". *Journal of Environmental Economics and Management*, in press (due 7/95).

Feather, Peter. "Sampling and Aggregation Issues in Random Utility Model Estimation". *American Journal of Agricultural Economics*, 76 (Nov. 1994), pp. 772-780.

Hanemann, W. Michael. "Valuing the Environment through Contingent Valuation". *Journal of Economic Perspectives*. 8 (Fall 1994).

Hansen, Leroy and Arne Hallam. "National Estimates of the Recreational Value of Streamflow". *Water Resources Research*. 27 (Feb. 1991), pp. 167-175.

Heimlich, Ralph. In *Designing Green Support Programs*. Henry A. Wallace Institute for Alternative Agriculture, Policy Studies Program Report No. 4, ed. Sara Lynch. Dec. 1994.

Hellerstein, Daniel. "Correcting for Bias When Average Values are Used to Compute Changes in Consumer Surplus". *Leisure Sciences*. 14 (1992),. pp. 337-345.

Johnson, R., E. Ekstrand, J.R. McKean, and K. John. "Economics of Wildlife and CRP" in *When Conservation Program Contracts Expire: The Policy Options*. Soil and Water Conservation Society, Ankeny, Iowa. (1994).

Kellogg, R., M. Maizel and D. Goss. *Agricultural Chemical Use and Ground Water Quality: Where are the Potential Problem Areas?* USDA, Economic Research Service. Dec. 1992.

Osborn, T. F. Llacuana, and M. Linsenbigler. *The Conservation Reserve Program: Enrollment Statistics for Signup Periods 1-11 and Fiscal Years 1990-92. USDA/ERS Statistical Bulletin No. 843*. (1992).

Palmquist, R.B. and L.E. Danielson. "A Hedonic Study of the Effects of Erosion Control and Drainage on Farmland Values", *American Journal of Agricultural Economics*. 71 (Feb.

1989), pp. 55-62.

Parsons, George and Michael Needleman. "Site Aggregation in a Random Utility Model of Recreation". *Land Economics*, 68, 418-433 (1992).

Portney, Paul (1994). "The Contingent Valuation Debate: Why Economists Should Care". *Journal of Economic Perspectives*. 8 (Fall 1994).

Ribaudo, Marc. "Estimating Changes in Recreational Fishing Participation from National Water Quality Policies". *Water Resources Research* 27 (1991), pp. 1757-1763.

Ribaudo, Marc, Daniel Colacicco, Linda Langner, Steven Piper, and Glenn Schaible. *Natural Resources and Users Benefit from the Conservation Reserve Program*. USDA, Economic Research Service Agricultural Economic Report No. 627 (1990).

U.S. Department of Agriculture, Economic Research Conservation Service (1994). *Agricultural Resources and Environmental Indicators*. Agricultural Handbook No. 705, ERS-NABS, Washington DC, Dec. 1994.

United States Department of Agriculture, Soil Conservation Service (81). "Farm Real Estate Values, Cash Rents, and Taxes, in *Land Resource Regions and Major Land Resource Areas of The United States*. Agriculture Handbook 296. Dec. 1981.

U.S. Department of Agriculture, Soil Conservation Service (1984). "*National Resources Inventory: A Guide for Users of 1982 NRI Data Files*. USDA/SCS, Washington D.C. and Iowa State Statistical Lab., Iowa State University, Ames, IA. (1984).

U.S. Department of Agriculture (1989). "*The Second RCA Appraisal: Soil, Water, and Related Resources on Nonfederal Land in the United States; Analysis of Conditions and*

Trends. USDA Miscellaneous Publication No. 1482, Washington D.C. (1990).

U.S. Department of Interior, Fish and Wildlife Service and U.S. Department of Commerce, Bureau of the Census (FHWAR). *1991 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation.* U.S. Government Printing Office, Washington D.C., (1993).

A Repeated Nested Logit Model of Fishing Participation, Site Choice, and Trip Duration

Heng Z. Chen, John P. Hoehn, Frank Lupi, and Theodore Tomasi

Department of Agricultural Economics
Michigan State University
East Lansing, MI 48824-1039

The research presented here has been partially supported by the Michigan Department of Natural Resources. The views expressed here are those of the authors and do not represent the views of the Michigan Department of Natural Resources. The results presented here are preliminary.

A Repeated Nested Logit Model of Fishing Participation, Site Choice, and Trip Duration

1. Introduction

When modelling recreational site choice, researchers are faced with many model specification issues. In this paper, we focus on how to treat recreation trips of different durations. The treatment of on-site time has received some attention in the travel cost literature. Trip duration and on-site time issues have been analyzed within a single site model (McConnell, Wilman). Within the context of multiple site models, the distinction among trips of different duration has received less attention.

In situations where there are many substitutes, recreation site choice models are increasingly based on the multinomial logit or the nested logit (NL). One reason for this popularity is that logit models can incorporate many substitute sites in a tractable way.[1] By including the option of not taking a trip at any choice occasion and by repeating this decision over the course of a season, the logit models can be used to explain seasonal participation and site choice (Morey, Shaw, and Rowe).

We introduce a model of recreational fishing site choice that places participation, site choice, and trip length decisions within a repeated nested logit framework. The model contains four levels of nests, a large number of alternative sites, and is estimated by Full Information Maximum Likelihood. The choice structure of the model makes the number of choice occasions endogenous and individual specific. By making trip length and the number of choice occasions endogenous, the decision of how to define choice occasion length is determined by the data instead of the researcher. This structure raises the question of how to predict the new number

---

[1] Morey provides a recent review of nested logit models with and without the participation option.

of trips of different durations and choices occasions following a change in site characteristics. We show that this calculation is straightforward.

This paper is organized as follows. Section 2 begins with a brief review of alternative treatments of trip length. Sections 2.1 and 2.2 present the behavioral model and demonstrate how the model can be used to predict total trips over a season. In section 3, the repeated nested logit is specified, and the data used to estimate the model is discussed. Estimation results are presented in Section 3.2. In section 3.3, we compare the model to repeated nested logits based on exogenous choice occasions and to a repeated logit where the distinction between trips of different durations is ignored. Section 4 concludes the paper.

2. The Model

When specifying a model of recreational site choice, there are several alternative ways of treating trips of different duration. One strategy is to limit the analysis to trips of one duration (Parsons and Needelman). With this approach, trips of different durations cannot serve as substitute activities. In addition, if the observed behavior involves trips of varying length, some data will necessarily be lost in this process.

Another approach is to combine trips into one generic type making no distinction between trips of different length (Feather, Hellerstein, and Tomasi). Treating all trips as if they are one duration can lead to misspecification. First, the choice set is likely to be misspecified because the feasible set will either be too big for day trips, too small for multiple day trips, or some combination of both. Second, when trip length is ignored, the value of on-site time will be misspecified.[2] Third, ignoring differences in trip duration forces each explanatory variable to have the same effect on the probability of taking a trip regardless of the actual trip length. In a model which permits different trip lengths, the equality of site quality parameters across trip lengths is a testable hypothesis. Moreover, in all of the approaches that only model one trip length, trips of different lengths can not serve as substitutes for one another.

Instead of ignoring trip durations, separate models can be constructed for each of the different trip lengths. The site choice models of Feather and of Jones and Sung follow this approach to modelling trips of different durations. However, this procedure does not make use of any joint information contained in the two decisions. In addition, separate models are not subject to any conditions which would make the decisions internally consistent. For example, the predicted number of days spent fishing could end up exceeding the season length.

There are several ways that the choice of trip length could be modelled in a repeated logit framework. One approach would be to select exogenous choice occasions which are long enough to allow each trip length to occur. For example, if desired trip lengths vary from one day to one week, choice occasions could be specified as one week long. Within each choice occasion, individuals could choose between the portfolio of possible trips which could occur at each trip length.[3] The data requirements of this approach are very demanding because one would need to know the dates of all trips to properly construct the portfolio associated with each choice occasion.

---

[2] A novel way to treat the trip cost is to impose some distance at which the on-site time of a trip changes. For example, Morey, Rowe, and Watson treat trips as one day if they were within State and multiple day if the were out of State. However, this approach imposes the restriction that multiple day trips can not be taken within some radius of an individuals origin.

[3] For example, Carson, Hanemann, and Wegge use weeks as the choice occasion, and they allow the choice between 1, 2, and 3 or more trips per week. One could expand this notion to allow combinations of trips of different durations.

Alternatively, one could impose that only one trip could occur per choice occasion. With exogenous choice occasions of fixed length, researchers would have to trim the data for any individuals who took more than one trip per choice occasion. The shorter is the length of the choice occasions, the less one would need to discard data because multiple trips occurred within a choice occasion. On the other hand, shorter choice occasions reduce the trip lengths which can be modelled. In this approach the decision on the number and length of choice occasions must be made by the researcher.

Consider an alternative model in which anglers make a repeated sequence of decisions to take a trip or not take a trip. To incorporate the choice among trip durations, suppose that each time a trip decisions made, the anglers compare the options of not going and going to each site at each trip duration. We impose the logical condition that anglers can not make a choice of taking a new trip while they are on a trip. Thus, if an angler takes a trip that lasts t days, the angler does not face another choice occasion for the next t-1 days. Upon returning from the trip, the angler repeats the choice process. Selecting a basic unit of time will determine the maximum possible number of choice occasions in a season. An individual's actual number of choice occasions will depend on the number of multiple day trips they choose. In this approach, choice occasions are determined by the data. Moreover, because the data does not need to be trimmed to fit an exogenous definition of choice occasions, the approach can make full use of all available data.

2.1 Model specification

In general, the recreational trip length is a continuous choice variable. As a practical matter, trip lengths can be partitioned into a number of discrete values such as $t = \{1,2,3,\ldots,T\}$. The basic discretization unit can be per day, per half day, or even per hour, depending on the specific application. On each choice occasion, an individual decides where to go and how long each trip will last. The joint probability of taking a trip to site j at length t is

$$Pr(j,t,p) = Pr(j|_{t,p})Pr(t|_{p})Pr(p)$$

where p indicates participation.

In our application, we will assume that the recreation trip decision is made on a daily basis. We impose the condition that the participation decision can not be made while on a trip. For each individual, the number of choice occasions in a season of S days is

$$N_{oc} = S - \sum_t N_t \cdot (t-1) \tag{3}$$

where N_t is the number of trips of length $t=(1,2,\ldots,T)$ in the season. $N = \sum_t N_t$ is the total number of trips in S days. The number of occasions without trips is given by $N_o = N_{oc} - N$. As a result, the likelihood function for an individual with N number of trips in the season is

$$L_i = \{\prod_N Pr(j,t,p)\}\{1 - \sum_{l,\tau} Pr(l,\tau,p)\}^{No}$$

where $\prod_N$ is the product over the N trips for the angler. Multiplying L_i over all individuals yields the sample likelihood function.

2.2 Trip prediction

The per occasion probability of taking a trip of length t is Pr(t,p). Thus, the predicted number of trips of length t in the season is:

$$N_t = N_{oc} Pr(t,p) \quad \text{for } t = 1,2,\ldots,T \tag{4}$$

On the other hand, (3) states that the number of choice occasions in the season is a function of the number of trips of various lengths. As a result, any change in site quality that affects choice probabilities will simultaneously induce a change in the number of trips and number of choice occasions. To predict the new number of trips, N_{oc} can be eliminated by substituting (3) into (4). This yields

$$N_t = \{S - \sum_t N_t \cdot (t\text{-}1)\} \Pr(t,p) \quad \text{for } t = 1,2,\ldots,T \tag{5}$$

The prediction is easily solved since we are left with T linear equations to solve for the T unknown trips N_t, for $t = 1,2,\ldots,T$.

3. Model Illustration

3.1 Data, Choice Structure and Probabilities

For illustration purposes, we apply the model to recreational fishing in Michigan. We use data from a small study of the fishing activities of Michigan residents over the 1993 fishing season. In this preliminary effort, we are not including individuals who did not fish at all during the season. After selecting individuals with a complete set of demographic variables, there are 195 anglers who took at least one trip. These individuals provided detailed information on site destination and trip length for 912 trips. In the sample data, 85% of trips are single day trips. The average total length of the multiple day trips is 3.42 days. Because of the small number of multiple day trips, we will model trip length as a choice between single day trips and a multiple day trip of 3.42 days.

Recreational fishing opportunities in Michigan include fishing at waters of the Great Lakes (GL), inland lakes (IL), and rivers and streams (RS). Great Lakes sites include Lake Michigan, Lake Superior, Lake Huron, Lake Erie, and connecting waters such as Lake St. Clair. Great Lakes sites are defined as stretches of GL shoreline within a county. 41 of Michigan's 83 counties have GL shoreline. Sites within the Inland lakes water bodies and the river and stream water bodies are also defined at the county level. For day trips we will allow anglers to choose GL sites, IL sites, or RS sites. Thus, there are 206 (41+83+83) potential alternative site/waterbody type combinations within each anglers day trip choice set.

Because the data for multiple day trips is more limited, nesting sites by water bodies in the manner of day trips results in some nests with few observations. Therefore, for the multiple day trips, we will nest sites by regions of the state. For multiple day trips, individuals can either choose to fish in counties in the Upper Peninsula (UP), counties in the Lower Peninsula bordering the Great Lakes (LPGL), and counties in the Lower Peninsula not bordering the Great Lakes (LPIL). For multiple day trips there are 83 alternatives within the choice set. Thus, individuals can have at most 290 options on each choice occasion, including 206 day trip alternatives, 83 multiple day trip alternatives, plus not going. In the application, each individuals choice set contains less than 290 alternatives because some sites are not within the feasible driving distance. Feasible sites for day (multiple day) trips are defined as the sites within the maximum observed driving distance for day (multiple day) trips.

The model contains four level of nests. One each for participation, trip length, waterbody/region, and site choices. See Figure 1. The joint choice probability of participation p at length t to waterbody/region w and site s is

$$\Pr(s,w,t,p) = \Pr(s|_{w,t,p}) \Pr(w|_{t,p}) \Pr(t|_{p}) \Pr(p).$$

Let the function $V_s(s|_{w,t,p})$ index the explanatory variables that vary across the choice of site s, conditional on waterbody/region, trip length, and participation.

$$\begin{aligned} V_s(s|_{w,t,p}) = \quad & \beta_0 Cost_{s,t} + \beta_1(1-D_t)Aoc_{s,w} + \beta_2(1-D_t)For_{s,w} + \beta_3 Cr_{s,w} + \beta_4 Acre_{s,w} + \\ & \beta_5 Mile_{s,w} + \beta_6 ILPAS_{s,w} + \beta_7 D_t Aoc_{s,w} + \beta_8 D_t For_{s,w} \end{aligned}$$

$Cost_{s,t}$ is the trip costs to site s of length t. Trip cost is defined as the sum of driving costs and time costs. Driving costs equal round trip driving distance multiplied by the AAA cost per mile for operating vehicles, \$0.38. Round trip distance is calculated using a matrix of driving distances between county centroids which was obtained from the Michigan Department of Transportation. Time costs equal the sum of driving and on-site time multiplied by the daily wage rate. The daily wage rate is defined as annual income divided by 365. Driving time is the driving distance (in miles) divided by an average speed of 50 miles per hour and translated into days by using 8 hours of driving time per day. On-site time is the sample average on-site time in days for single and multiple day trips.

The site quality information was obtained from the Michigan Department of Natural Resources. The site quality variables are defined as follows. $Aoc_{s,w} = 1$, if site s of waterbody w is designated by the International Joint Commission as a Great Lakes Area of Concern due to contamination, 0 otherwise. $For_{s,w}$ is percentage of the county that is covered by forest. $Cr_{s,w}$ is average seasonal catch rate of Chinook Salmon at Great Lakes sites and only enters for day trip sites in the GL nest. $Acre_{s,w}$ is the acreage of inland lakes per county and only enters for day trip sites in the IL nest. $Mile_{s,w}$ is miles of top quality streams and tributary streams and only enters for day trip sites in the RS nest. $ILPAS_{s,w}$ is the number of inland lake public access points within a county and only enters for day trip sites in the IL nest. $D_t=0$ if the trip is less than or equal to one day; $D_t=1$, if it is a multiple day trip. Through the interaction term, D_t, the coefficients of Aoc and For can take different values for the different trip lengths.

If the conditional probabilities are specified as Nested Logit, we have:

$$Pr(s|_{w,t,p}) = \exp(V_s(s|_{w,t,p})) \;/\; [\textstyle\sum_s \exp(V_s(s|_{w,t,p}))] \tag{6}$$

where $\sum_s$ means the sum over the feasible sites, given w and t. The inclusive value of taking a trip to waterbody/region w at length t is $IV(w,t,p) = \ln[\sum_s \exp(V_s(s|_{w,t,p}))]$.

On the third level, conditional on choosing trips of a single day, the fishing choices are GL, IL, and RS. To adjust for difference across the three water bodies, two dummy variables are used; GL=1 if fishing is at the GL, 0, otherwise, and IL=1 if fishing is at the IL, IL=0, otherwise. The conditional probability of visiting w for the single day trip is

$$\Pr(w|_{d,p}) = \exp(V^{d}_{w}(w|_{d,p})) \,/\, [\textstyle\sum_{w} \exp(V^{d}_{w}(w|_{d,p}))] \tag{7}$$

where d represents the day trip length and $V^{d}_{w}(w|_{d,p}) = \gamma_1 GL + \gamma_2 IL + \gamma_3 IV(w,d,p)$. Thus, the inclusive of taking a day trip is $IV^{d}(d,p) = \ln[\sum_{w} \exp(V^{d}_{w}(w|_{d,p}))]$.

For multiple day trips, the choices are UP, LPGL, or LPIL. Two dummy variables are used to account for differences among the regions: UP = 1 if the trip is to the UP, 0, otherwise; and LPGL=1, if the trip is to a county bordering the GL of the LP, 0, otherwise. The probability of choosing w, conditional on participation and taking the multiple day trip, is

$$\Pr(w|_{md,p}) = \exp(V^{md}_{w}(w|_{md,p})) \,/\, [\textstyle\sum_{w} \exp(V^{md}_{w}(w|_{md,p}))] \tag{7'}$$

where md represents multiple day and $V^{md}_{w}(w|_{md,p}) = \gamma_4\, UP + \gamma_5\, LPGL + \gamma_6\, IV(w,md,p)$. The inclusive value of taking a multiple day trip is $IV^{md}(md,p) = \ln[\sum_{w} \exp(V^{md}_{w}(w|_{md,p}))]$.

On the second level, the probability of taking a trip of length t, conditional on participation, is

$$\Pr(t|_{p}) = \exp(V_t(t|_{p})) \,/\, [\textstyle\sum_{\tau} \exp(V_t(\tau|_{p}))] \tag{8}$$

where $V_t(t|_{p}) = \delta_0\, D_t + \delta_1\, IV^{t}(t,p)$. The trip's inclusive value is $IV(p) = \ln[\sum_{\tau} \exp(V_t(\tau|_{p}))]$.

On the top level of nesting, not participating in fishing or "staying home" is indexed by $V(h) = \eta_0\, D_h + \eta_1\, EDU$, where EDU =1 if the individual's educational status includes some college, 0, otherwise. $D_h=1$ if no trip is observed at the occasion, 0 otherwise. The index for taking a trip is $V(p) = \eta_2 IV(p)$, as compared with staying home $V(h)$. Thus, on any choice occasion the probability of taking a trip is

$$Pr(p) = \exp(V(p)) / [\exp(V(h)) + \exp(V(p))]. \tag{9}$$

3.2 Estimation Results

Three models are estimated to investigate the importance of distinguishing the trip lengths and the definition of choice occasions. The first model is based on the endogenous choice occasions presented above. The second model combines all trip types into one generic length, and the third model employs an exogenous choice occasion with one trip per occasion. All the models are estimated using Full Information Maximum Likelihood (FIML).

Model 1. Model 1 (M1) uses the endogenous choice occasion definition (3) and the data set of 195 anglers with 912 trips. Trips that last one day or less than one day are classified as single day trips while trips that last more than one day are multiple day trips. The estimation results are given in Table M1 in the Appendix. Recall that the parameters are given by β's for site quality variables, by γ's for variables in waterbody/region level, by δ's for variables in trip length choice level, and by η's for the participation level. The estimation results show that most of the parameters have the expected sign and most are also significant. Specifically, the parameter estimates of all inclusive values γ_3, γ_6, δ_1, and η_2 are within (0,1) interval.

To examine the usefulness of the nest structure, we estimated the independent multinomial logit model by restricting the parameters of IV variables γ_3, γ_6, δ_1, and η_2 to one. The log likelihood value of the independent logit model is -8319.057. The likelihood ratio test statistic is 2110 with four degrees of freedom. The test indicates that the nest structure considered in this model is a significant improvement over an un-nested version.

To check how well the model corresponds to the actual observations, we use (5) to predict the trips at the baseline quality level (q_0). As presented in Table 1, the predictions coincide with the observations quite well at the baseline q_0. Table 1 also presents the predicted change in trips after the removal of the Aoc status for all affected counties (q_1). The negative Aoc parameters β_1 and β_7 (see Table M1 in the appendix) for the day and multiple day trips imply that the value of the index of explanatory variables for both trip types will increase with the policy. After the policy change, single day trips increase 1.79, while the multiple day trips decrease by 0.14. The total trip change for the 195 anglers is 1.65, or 16,800 trips statewide[4].

Table 1: M1 trips at q_0 (the baseline) and at q_1 (without Aoc):

	Single day	Multiple day	Total
Observed trips at q_0 (195 anglers)	780	132	912
Predicted trips at q_0 (195 anglers)	780.11	131.76	911.87
Trip change due to $q_0 \Rightarrow q_1$ (195 anglers)	1.79	-0.14	1.65
Trip change due to 1.57 cent/mile cost increase at q_1	-2.34	0.69	-1.65
Trip change due to $q_0 \Rightarrow q_1$ (statewide)	N/A	N/A	16,800

[4] We first predict the trip demand in the full season of 232 days (from April to mid-November) for each of the 195 anglers. Then, as a crude approximation, the statewide trip demand is calculated as the product of per angler's demand and 1,436,800, the number of resident anglers in Michigan in 1991 (National Survey of Fishing, Hunting, and Wildlife-Associated Recreation, 1994).

In this example, some day trips are substituted for multiple day trips, i.e., the Aoc policy leads to an increase in single day trips but a decrease in multiple day trips. This is because the probability of taking a trip of length t is $Pr(t,p) = Pr(t|_p)Pr(p)$. Although $Pr(p)$ for participation increases due to the positive parameters of the inclusive values (γ_3, γ_6, δ_1, and η_2), the conditional probability $Pr(md|_p)$ for taking a multiple day trip actually decreases, while $Pr(d|_p)$ for taking a day trip increases. The values of $Pr(d|p))$ and $Pr(md|p))$ are substitutes according to $\Sigma_t (Pr(t|p)) = 1$ as can be seen from (8). Since the decrease in $Pr(md|_p)$ for the multiple day trip is not offset by the increase in $Pr(p)$, the joint probability for multiple day trips $Pr(md,p) = Pr(md|_p)Pr(p)$ decreases, while the joint probability $Pr(d,p)$ for the single day trip increases. This suggests that as site quality improves, anglers may not take more trips of all lengths though this is possible. Overall, the quantity of trips does increase as site quality increases.

To further evaluate the implications of the Aoc policy, we solved for the price change that would leave total trips at the baseline level given that the Aoc policy was implemented. This comparison is meant to illustrate the importance of the policy change even though it is not a welfare measure. Once again, the single day trips and multiple day trips respond to the trip cost policy differently. We found that in the post policy state (q_1) a 1.57 cent per mile increase in driving cost would reduce the total trips back to the trip level corresponding to the baseline level (q_0). The single day trips decrease by 2.34 as compared with trips in the post-policy state q_1, or 0.55 as compared with trips in the baseline q_0. On the other hand, the multiple day trips increase by 0.69 over the post-policy trip demand at q_1, or 0.55 as compared with the trip demand at the baseline q_0. The reason for change in the mix of day and multiple day trips is that the driving cost increase of 1.57 cent per mile increases the relative cost of taking a day trip as compared with the multiple day trips since the costs of a trip is the sum of the driving cost and time cost. As a result, single day trips decrease while multiple day trips increase.

3.3 Model Comparisons

Model 2. To investigate the importance of separating trips based on the trip lengths, we estimated a second model, Model 2 (M2). In M2 all trips are treated as if there were no difference in trip length. That is, all trips last 1.35 days, the average trip length from the sample. 1.54 days is used as the length of the exogenous choice occasion because it is the minimum time interval in which no more than one trip is observed in the sample data set. The feasible choice set for M2 is defined as the set of sites that are within the maximum observed driving distance for the sample. Thus, the choice set mirrors the multiple day trip choice set of model 1. The nesting structure follows the nesting structure used in the day trip segment of Model 1, though there is no level for the trip duration. Likewise, there is no dummy variable D_t to separate the trips according to their lengths.

The estimation results are presented in Table M2. As in M1, the β's are the parameters for the variables that vary at the site choice level. The γ's are for the variables of waterbody level, and the η's are for the participation level. Again, all the parameters have the expected sign in M2. γ_3 and η_2 for the inclusive values are also significant and are between 0 and 1. However, by comparing M1 and M2, it is noticed that the parameter on the travel cost per hundred dollars, β_0, is different between the two models. β_0=4.36 for M1, while β_0=3.28 for M2. Although the predicted trips from both M1 and M2 correspond to the observed trips at baseline quite well, the predicted change in trips due to the removal of the Aoc statewide is different due to the structural differences between the two models. See Table 2.

Table 2: M2 trips at q_0 (the baseline) and q_1 (without Aoc):

Observed trips at q_0 (195 anglers)	Predicted trips at q_0 (195 anglers)	Trip change due to $q_0 \Rightarrow q_1$ (195 anglers)	Trip change due to 1.12 cent/mile cost increase at q_1 (195 anglers)	Trip change due to $q_0 \Rightarrow q_1$ (statewide)
912	911.99	1.06	-1.06	7,900

In the post-policy state, q_1, the predicted change in trips from M2 using the sample of 195 anglers is 1.06. At the statewide level, by using the previous extrapolation method, the trips increase by 7,900. The change in trips is less than half that of M1. Furthermore, we found that from the post-policy state, an increase in travel cost of 1.12 cent per mile would reduce trips to the original baseline level. This compares with 1.57 cent per mile for M1.

Model 3. We also compared the endogenous choice occasion model M1 with an exogenous choice occasion model in which the difference in trip lengths is recognized. To compare the impacts of choice occasion definitions, Model 3 (M3) uses 5 days as the length of a choice occasion. To estimate the model with exogenous choice occasions some trips need to be trimmed away from the avid anglers so that the exogenous choice occasion definition can be employed. In particular, when we define the length of the exogenous occasion be 5 days, 42 trips will have to be trimmed away from ten avid anglers who took more than one trip every five days. We randomly chose the trips that would need to be trimmed to avoid any systematic bias in the way we trimmed the data. One angler is trimmed away because we no longer know the details of where the angler went on the randomly selected trips. We are left with a trimmed data set that consists of 194 anglers with a total of 870 trips, including 740 single day trips.

As in M1, trips are separated into single day trips and multiple day trips. The feasible choice set is defined differently for the different trip lengths by using the maximum observed driving distance in each trip length category. The parameter estimates are provided in Table M3. It is noticed that by comparing M1 with M3, the participation parameters η's differ between the two models. For example, η_0 for the dummy of staying home decreases from 3.51 for M1 to 1.97 for M3 due to the change in the length of the choice occasion. In M1, education is significant in explaining the participation, while this is not the case for M3. The parameters on travel cost also differ. The policy impacts for M3 are reported in Table 3.

Table 3: M3 trips at q_0 (the baseline) and q_1 (without Aoc):

	Single day	Multiple day	Total
Observed trips at q_0 (194 anglers)	740	130	870
Predicted trips at q_0 (194 anglers)	739.44	130.56	870
Trip change due to $q_0 \Rightarrow q_1$ (194 anglers)	2.32	-0.44	1.92
Trip change due to 2.10 cent/mile cost increase at q_1	-2.86	0.97	-1.92
Trip change due to $q_0 \Rightarrow q_1$ (statewide)	N/A	N/A	16,400

The trip predictions once again fit the baseline quite well. The trip change for the sample of 194 anglers is 1.92, as compared with 1.65 trips for the 195 anglers in M1. When extended to the whole season and extrapolated statewide, the trip change is 16,400 for M3, as compared with 16,800 from M1. Another difference between M1 and M3 is that the increase in travel costs which returns total trips to the pre-policy level is 2.10 cent per mile, as compared with

1.57 cent per mile in M1. These examples illustrate the sensitivity to the treatment of trip length and the definition of choice occasions.

4. Conclusions

Previous researchers have used repeated logit models to endogenize the decision of which site to visit and whether or not to participate (Morey, Shaw, and Rowe, and Morey, Rowe, and Watson, among others). We extended the research by developing a repeated logit model that also endogenizes the choice of trip length. The choice of trip length is a feature that has been overlooked by many previous research efforts (two exceptions are Jones and Sung, and Feather). By explicitly modelling the trade-off among trip lengths, participation, and site choice, the model allows substitution among trip lengths in addition to site choice and trip participation. It can accommodate the interesting possibility that trips of some lengths could decrease while trips of others lengths could increase as a result of improvements in site characteristics.

We estimated three models of the demand for recreational fishing trips in Michigan. The models are all repeated, three or four level nested logits which were estimated using Full Information Maximum Likelihood methods on a large choice set. Because of the FIML estimation, the parameter estimates for each nest are efficient. Likelihood ratio test statistics calculated from nested and non-nested versions of M1 suggests that the nest structure employed in this paper is highly significant.

Further comparisons among the three models illustrated that the treatments of different trip lengths and the choice occasions can lead to different policy implications, as presented in Tables 1, 2, 3 of the previous section. It is noted that since the model formulation is based on the highly selective sample and the limited site quality information, the model results may not well represent the behavior of all Michigan anglers. As such, our application is only meant to illustrate the feasibility of the approach. There are many remaining research issues to be addressed, such as the elementary site definition and alternative choice structures. With more sample observations, the choice of trip lengths would not be limited to the single day or the multiple day distinctions employed here.

Appendix

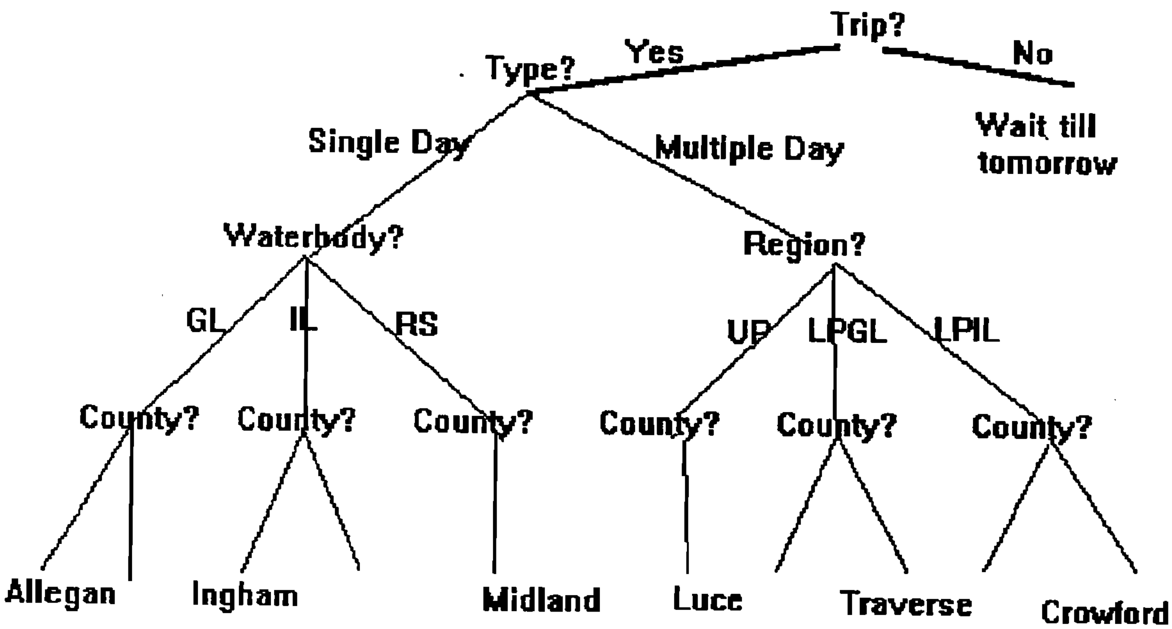


Figure 1 Nest Structure

Table M1: Estimation results of the endogenous choice occasion model

Endogenous Choice Occasion Models	M1 (195 anglers) LL = -7264.0646
Variables	Estimates (t-Stat)
β_0 Cost/100	-4.3638 (-35.3177)
β_1 Aoc(d)	-0.3127 (-1.7281)
β_2 Forest(d)	0.9850 (3.3721)
β_3 Cr(d)	1.5719 (0.5757)
β_4 Acres(d)	0.2246 (4.0759)
β_5 Miles(d)	0.4176 (2.1089)
β_6 ILPAS(d)	0.0395 (7.8927)
β_7 Aoc(md)	-2.5348 (-4.8315)
β_8 Forest(md)	12.8833 (20.3626)
γ_1 GL(wbdy)	0.9652 (4.0248)
γ_2 IL(wbdy)	0.2226 (1.4503)
γ_3 IV(wbdy)	0.8963 (5.7861)
γ_4 UP(regn)	-0.5374 (-1.5728)
γ_5 LPGL(regn)	-0.2299 (-1.1004)
γ_6 IV(regn)	0.1062 (3.5348)
δ_0 D_t	-2.3100 (-13.6212)
δ_1 IV(t)	0.3512 (3.0656)
η_2 IV(p)	0.1500 (2.4276)
η_0 D_h	3.5054 (83.0420)
η_1 Edu	-0.2677 (-3.0546)

Table M2: Estimation results of exogenous choice occasion of 1.54 days and no distinction among trip lengths

Exogenous Choice Occasion Model	M2: OC=1.54 LL = -6946.2275 (195 anglers)
Variables	Estimates (t-Stat)
β_0 Cost/100	-3.2764 (-34.7237)
β_1 Aoc	-0.2614 (-1.4979)
β_2 Forest	2.7228 (11.2581)
β_3 Cr	3.3916 (1.2424)
β_4 Acres	0.1779 (3.6233)
β_5 Miles	0.3294 (1.7045)
β_6 ILPAS	0.0336 (7.1611)
γ_1 GL	0.6528 (2.9827)
γ_2 IL	0.4237 (3.4395)
γ_3 IV(w)	0.6550 (4.3912)
η_2 IV(p)	0.0613 (2.6833)
η_0 D_h	3.1182 (72.486)
η_1 Edu	-0.2746 (-3.1205)

Table M3: Estimation results with exogenous choice occasion of 5 days

Exogenous Choice Occasion Models	M3 (OC = 5 days) LL =-5690.4729 (194 anglers)
Variables	Estimates (t-Stat)
β_0 Cost/100	-4.2892 (-34.4272)
β_1 Aoc(d)	-0.4015 (-2.0856)
β_2 Forest(d)	1.0291 (3.4004)
β_3 Cr(d)	2.0039 (0.7083)
β_4 Acres(d)	0.1897 (3.2264)
β_5 Miles(d)	0.4726 (2.3244)
β_6 ILPAS(d)	0.0411 (8.0266)
β_7 Aoc(md)	-2.7645 (-5.0712)
β_8 Forest(md)	12.6114 (19.9909)
γ_1 GL(wbdy)	0.9781 (4.0394)
γ_2 IL(wbdy)	0.2906 (1.9024)
γ_3 IV(wbdy)	0.8560 (5.5866)
γ_4 UP(regn)	-0.6176 (-1.8026)
γ_5 LPGL(regn)	-0.2325 (-1.1135)
γ_6 IV(regn)	0.1064 (3.5218)
δ_0 D_t	-2.1924 (-13.1082)
δ_1 IV(t)	0.3739 (2.9747)
η_2 IV(p)	0.1548 (2.3495)
η_0 D_h	1.9742 (44.2833)
η_1 Edu	-0.0439 (-0.4588)

References

Carson, Richard T., W. Michael Hanemann, and Thomas Wegge, Southcenteral Alaska Sport Fishing Economic Study, *Mimeo*, November, 1987.

Feather, Peter M., Valuing Water Quality Using Discrete Choice Models, PhD Dissertation, University of Minnesota, 1992.

Feather, Peter M., Daniel M. Hellerstein, and Theodore Tomasi, A Discrete-Count Model of Recreational Demand, *Journal of Environmental Economics and Management*, forthcoming 1995.

Jones, Carol A., and Yusen D. Sung, Valuation of Environmental Quality at Michigan Recreational Fishing Sites: Methodological Issues and Policy Applications, Report to the EPA, 1993.

McConnell, K.E., On-Site Time in the Demand for Recreation, *American Journal of Agricultural Economics*, 74(4):918-925, 1992.

Morey, Edward R., Two RUMs unCloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice, *W-133 Benefit and Costs Transfer in Natural Resource Planning*, Seventh Interim Report, Western Regional Research Publication, June 1994.

Morey, Edward R., W. Douglass Shaw and Robert D. Rowe, A Discrete-Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data Are Not Available, *Journal of Environmental Economics and Management*, Vol. 20:181-201, 1991.

Parsons, George R., and Michael S. Needelman, Site Aggregation in a Random Utility Model of Recreation, *Land Economics*, 68: 418-33, 1992.

Wilman, Elizabeth A., A Simple Repackaging model of Recreational Choices, *American Journal of Agricultural Economics*, 69(3):603-612, 1987.

Limiting Program Contradictions Through Analysis of Utility Theoretic Demand Systems

Abstract

Managing programs and program elements one-at-a time produces contradictory management plans. An important step towards reducing program contradictions is information on the interdependence of marginal values across program elements. This paper presents a utility-theoretic partial demand system that incorporates several dimensions of quality for a system of New Mexico fishing reservoirs and streams. The model permits estimating the benefits of competing policy proposals in which benefits are sensitive to the context in which the proposals occur. Results illustrate that existing conditions at all sites in the system influence the value of a policy that alters conditions at any site.

Frank A. Ward
Department of Agricultural Economics and Agricultural Business
New Mexico State University
Las Cruces, NM 88003

Richard A. Cole
Department of Fishery and Wildlife Sciences
New Mexico State University
Las Cruces, NM 88003

March 1995

This paper was presented at the Western Regional Research Project "Benefits and Costs Transfer in Natural Resource Planning," (W-133) Monterey, California, March 13, 1995. Funding was provided by the New Mexico State University Agricultural Experiment Station, New Mexico Department of Game and Fish, and New Mexico Water Resources Research Institute.

Limiting Program Contradictions Through Analysis of Utility Theoretic Demand Systems

BACKGROUND

Natural resource managers typically pursue proposals independently. They focus on single proposals and seldom consider what managers in the same or other agencies may be considering. Optimizing one program at a time assumes it is independent of other programs; however, programs are interdependent.

The benefit of a planned improvement at a given site depends on plans for all sites in the system. For example, the recreation benefit of a program element that limits water drawdowns at one reservoir depends on programs that set lake levels and facilities at substitute reservoirs.

Managing sites and site quality elements one at a time produces contradictory management plans (Lave, 1984). The total value of the package of policy elements is not the value of the individual parts added up (Lichtenberg and Zilberman, 1986; Hoehn and Randall, 1989). Independent valuation of interdependent elements produces a mix of policy elements that fails to maximize total benefits: Complementary program elements are underfunded and substitute elements are overfunded.

Economically efficient plans identify the mix of program elements over time and space that maximize benefits for available resources. These optimal plans are conditioned by demographics, social values, and site conditions. Managers who wish to pursue optimal programs need information on how the planned level of each program element within a system influence the incremental benefit from altering any program element.

An important challenge facing benefit-cost analysts is specifying and estimating models that measure the benefits of policies that reconfigure characteristics of a system of sites that contain interdependent elements of value. Viewed incrementally, the challenge is developing models that measure marginal benefits of improving any site as a function of conditions at all sites.

This paper presents a utility-theoretic partial demand system that incorporates several dimensions of quality for a system of New Mexico fishing reservoirs and streams. The model permits managers to estimate the benefits of competing policy proposals, in which benefits are sensitive to the proposals' contexts. Results illustrate that existing conditions at all sites in the system influence the value of a policy that alters conditions at any site.

EMPIRICAL MODEL

Data

Data on fishing demands, fishing catch, harvest rates, and average size of caught fish were obtained from statewide, monthly telephone surveys of New Mexico anglers conducted during 1988-1989. Other dimensions of site quality were obtained from various sources, including the regional New Mexico fisheries managers. Other sources for site quality data included Forest Service, Bureau of Land Management, and existing inventories of state park facilities. Where possible, published data on site facilities were verified by telephone followups.

The price of fishing was specified as the sum of travel cost plus the value of travel time (valued at one-half the average wage rate) plus all relevant site entry fees. Various demographic data were obtained from standard census sources. The complete data set consisted of 9504 observations, including observations on each of 132 sites, 9 zones of origin, and 8 calendar seasons for the period 1988-1989.

Model Performance Criteria

A model that effectively evaluates competing policies and programs implemented at site systems should be based on a system of demands, so that substitution across a changing opportunity set due to policy decisions can be accounted for. A demand system consistent with choice theory assures the budget constraint is enforced, thus allowing the benefits of competing policies to be grounded in choices and constraints actually facing consumers.

The model should adapt to a large number of spatially separated sites (or goods), in which each site consists of several measurable characteristics. Prices and qualities at all sites in the market area should affect each site's demand. The model should also account for demographic characteristics of visitors from the various sampling units (e.g. zones of origin), so estimates of benefits are sensitive to the spatial and temporal distribution of how preferences vary by visitors.

To be consistent with the principle of diminishing marginal returns, policy elements that produce any quality improvement at any site should produce diminishing marginal visits and marginal benefits. Finally, we looked for a system that was linearizable in the parameters to avoid estimation based on tedious iterative regression procedures.

Model Specification

A review was conducted of several widely used demand systems consistent with choice theory. Several of these are reviewed in Pollak and Wales (1992), including the Linear

Expenditure System (LES), the translog system, the Cobb-Douglas, and Constant Elasticity of Substitution (CES). For the present model, we wanted to select a system that required few estimated parameters for own- and cross-site price effects so more parameters could be reserved for quality effects and large numbers of sites could be accommodated. The LES requires at least n estimated price effects parameters for an n site demand system; the translog requires $(n/2)(n+1)$. Because our system had more than 100 sites, both these specifications were rejected.

Because of its simplicity, we conducted several policy simulations using the Cobb-Douglas utility function demand system. However, the Cobb-Douglas was rejected because its unitary elasticity of substitution effectively ignores substitute site prices. That is, an increase in entry fee or travel cost at any given site has no effect on the demand for any of the other sites in the system.

Extensive graphical, numerical, and mathematical testing of a constant elasticity on substitution preference ordering (CES) augmented by site quality characteristics was conducted. This algebraic form best satisfied the set of model performance criteria described above:

Indices

The indices for the model are defined as:

i = fishing site = 132 New Mexico fishing reservoirs and streams

j = one of 9 New Mexico zones of origin

k = 1 of 18 site quality indicators shown in Table 1

t = Calendar season = winter, spring, summer, fall, 1988-1989.

L = 1 of 7 demographic variables shown in Table 2

Variables

Model variables are defined as:

X_{ijt} = predicted trips to the ith site from jth zone for the tth season.

P_{ijt} = (round trip miles x travel cost/mile + entry fee); travel cost includes an opportunity cost of time valued at 1/2 the average hourly wage for the zone of origin.

M_{jt} = total fishing expenditure per angler from jth zone of origin, tth season = $\sum_i P_{ijt} X_{ijt}$.

Q_{ikt} = individual site facility variables of the ith site kth quality indicator for the tth season.

G_{ijht} = geographical quality variables of the ith site hth quality indicator for the jth zone and tth season.

D_{Ljt} = individual Lth demographic variable of jth zone for the tth season, $0 < D_{Ljt} < 1$ (Table 2).

Z_{jt} = composite demographic zone index of jth zone, tth season, a combination of zone demographic variables shown in Table 2, where $0 < Z_{jt} < 1$.

β_{ijt} = function of individual quality indicators and zone index that enters angler preference ordering for the ith site, jth zone, for the tth season.

A_{it} = administratively closed site variable for the ith closed site, for the tth season. $A_{it}=1$ when site is open. $A_{it}=0$ when site is closed.

The Model

The direct fishing satisfaction (utility) index is specified as the following quality-augmented CES function:

(1) $$U_{jt} = [\sum_i (\beta_{ijt} X_{ijt})^{\rho}]^{1/\rho}, \qquad \rho \le 1 .$$

where the term β_{ijt} is related to quality and demographics as shown below.

A system of demands predicts fishing trips per jth zone angler to each ith site for the tth season. Angler demands are based on the assumption that the angler acts as if maximizing the satisfaction index (1), subject to the limited fishing budget, M_{jt}. The result of that maximization produces the following system of demands:

$$(2) \qquad X_{ijt} = \frac{M_{jt}\, \beta_{ijt}^{\sigma-1}\, P_{ijt}^{-\sigma}}{\sum_i \beta_{ijt}^{\sigma-1}\, P_{ijt}^{1-\sigma}}, \qquad \sigma > 0 ,$$

where,

$$(3) \qquad \beta_{ijt} = A_{it} \prod_k (Q_{ikt})^{\gamma_k Z_{jt}} \prod_h (G_{ijht})^{\gamma_h Z_{jt}} .$$

That is β_{ijt} depends on site quality factors and geographical factors, both of which interact with demographic variables. In equation (3), Z_{jt}, the zone index, is defined as:

$$(4) \qquad Z_{jt} = \prod_L (D_{Ljt})^{\alpha_L}, \qquad 0 < D_{Ljt} \le 1 ,$$

as shown in Table 2. That is an overall demographic index depends on several demographic elements.

The demand system X_{ijt} defined in (2) is derived from the assumption that the angler maximizes (1) subject to the fishing expenditure constraint:

$$(5) \qquad M_{jt} = \sum_i P_{ijt} X_{ijt} .$$

The indirect fishing satisfaction function is obtained by substituting the equilibrium demand system (2) into the direct satisfaction index (1). It results in the following:

$$(6) \qquad U_{jt}^{\circ} = M_{jt} \,/\, C_{jt}^{\circ},$$

where

$$(7) \qquad C_{jt}^{\circ} = (\sum_{i} \beta_{ijt}^{\circ\ \sigma-1} P_{ijt}^{\circ\ 1-\sigma})^{1/(1-\sigma)}.$$

The (°) superscript in (6) and (7) indicate pre-policy values of variables.

The indirect satisfaction function can be inverted to solve for the expenditure function. It explains the angler's minimum total fishing expenditure required under <u>post-policy</u> levels of P_{ij}, Q_{ikt}, to achieve the same fishing satisfaction index as under <u>pre-policy</u> satisfaction, U_{jt}°. The expenditure function is obtained by inverting the indirect satisfaction index (6) and solving for expenditure as a function of satisfaction. This results in

$$(8) \qquad E_{jt}^{\circ} = U_{jt}^{\circ} (\sum_{i} \beta_{ijt}^{\ \sigma-1} P_{ijt}^{\ 1-\sigma})^{1/(1-\sigma)},$$

where E_{jt}° is interpreted as the minimum expenditure to sustain pre-policy angler welfare under post-policy conditions.

Benefits per angler measure the angler's welfare change, relative to pre-policy conditions, resulting from a new policy. We use the compensating variation (*CV*) welfare indicator, measured as:

(9) $$CV_{jt} = M_{jt} - E^{\circ}_{jt}.$$

Total benefits from a policy change at the *j*th zone of the *t*th season are found by multiplying per capita benefits by angler population as:

(10) $$Benefits_{jt} = (CV_{jt})(POP_{jt})$$

where POP_{jt} is the estimated angler population in the *j*th zone in the *t*th season. Total statewide benefits from a policy change are found by summing (10) over zones of angler origin and over relevant time periods:

(11) $$Benefits = \sum_{j}\sum_{t} Benefits_{jt}.$$

Marginal benefit per angler of an improvement in a site quality indicator is calculated as the change in CV_{jt} in (9) with respect to a single quality indicator, Q_{ikt}. The marginal benefit per angler of the *k*th quality indicator at the *i*th site, *j*th zone in *t*th season is specified in the following way:

(12) $$MB_{ijkt} = \frac{\partial CV_{jt}}{\partial Q_{ikt}},$$

where the term, $\frac{\partial CV_{jt}}{\partial Q_{ikt}}$, is the change in the total benefits with respect to an individual quality indicator in Table 1.

The term $\frac{\partial CV_{jt}}{\partial Q_{ikt}}$ is obtained by differentiating (9) with respect to Q_{ikt}, which is:

$$(13) \quad MB_{ijkt} = M_{jt} P_{ijt}^{1-\sigma} \gamma_k Z_{jt} Q_{ikt}^{-1} (\beta_{ijt})^{\sigma-1} \frac{(C_{jt})^{\sigma}}{C_{jt}^{\circ}},$$

where C_{jt} is defined in (7).

Aggregate marginal benefits over all anglers at the *i*th site, *j*th zone, and tth season are found by multiplying per capita marginal benefits in (13) by angler population:

$$(14) \quad MBA_{ijkt} = (MBQ_{ijkt})(POP_{jt}),$$

where POP_{jt} is the estimated angler population in the *j*th zone in the *t*th season. Total annual statewide marginal benefits per quality indicator per site for a given time interval (e.g., a water year) from a policy change are found by summing (14) over zones and the relevant time periods:

$$(15) \quad MBT_{ik} = \sum_{j} \sum_{t} MBQ_{ijkt}.$$

Examination of the demand system (2) shows that there is no separate set of parameters required for any given site. This specification allows the greatest generalizability and transferability to unstudied areas. Demand for any given site depends only on price and quality of that site and prices and qualities of all substitutes in the market area. Sites are differentiated only by their prices and characteristics. Nearby or higher quality substitutes have stronger effects on site visitation and marginal benefits of a quality improvement than more distant or lower quality substitutes. More sites provide more observations for fitting the parameters. However, one parameter must be estimated for each extra quality characteristic specified.

Equation (2) also shows that prices and qualities of each site in the system affect each site's demand. The denominator of (2) accounts for effects of price or quality changes at all sites, while the numerator accounts for similar effects at the "own" site. Demographic characteristics are captured in the exponent term in (3).

Context sensitivity is an important feature of this CES demand system. Inspection of the utility, demand, expenditure, and marginal benefits functions show that all these functions are sensitive to the context in which the policy change occurs. Substituting (13) into (3) shows that the marginal benefits of any quality improvement at any site depends on all site prices, all site characteristics, and all demographic features of visitors in the market area. For example, the benefits of increasing the average size of caught fish by 10 percent varies considerably according to the existing condition of the site, location of the site, demographic characteristics of market area anglers, availability of substitutes, and hydrological conditions. The economic value of a site improvement program depends on the context in which it occurs.

One added advantage of fitting a demand system known to be consistent with choice theory, such as the CES with quality, is that complex consumer surplus mathematical integrations are not required to perform welfare analysis. Once the demand system coefficients are estimated, they can be inserted directly into the expenditure function (8). The expenditure function permits non-marginal welfare analysis, i.e. computation of the benefits of program elements that implement a series of finite price or quality changes at one or more sites. Inserting the coefficients directly into the marginal benefits function (12) permits marginal welfare analysis, in which the change in benefits of any one unit program element change can be calculated.

Model Estimation

Equation (2) is transformed in several steps to permit estimation with linear regression methods. The dependent variable in (2) is transformed from quantity demanded to the log ratio of expenditure at the given site to expenditure at a base site. The derivation begins by defining the budget share allocated to a given site as:

$$(16) \quad f_{ijt} = \frac{X_{ijt}P_{ijt}}{M_{jt}} = \frac{\beta_{ijt}^{\sigma-1}P_{ijt}^{1-\sigma}}{\sum_i \beta_{ijt}^{\sigma-1}P_{ijt}^{1-\sigma}}.$$

Dividing both sides of (16) by f_{Ijt}, the budget share at some particular Ith site, causes the denominator of (16) to drop out. This results in

$$(17) \quad \left[\frac{f_{ijt}}{f_{Ijt}}\right] = \left[\frac{\beta_{ijt}^{\sigma-1}P_{ijt}^{1-\sigma}}{\beta_{Ijt}^{\sigma-1}P_{Ijt}^{1-\sigma}}\right] = \left[\frac{\beta_{ij}}{\beta_{Ij}}\right]^{\sigma-1}\left[\frac{P_{ijt}}{P_{Ijt}}\right]^{1-\sigma}.$$

Taking natural logs of both sides of (17) produces

$$(18)\quad \ln\left[\frac{f_{ijt}}{f_{Ijt}}\right] = (\sigma-1)\ln\left[\frac{\beta_{ijt}}{\beta_{Ijt}}\right] + (1-\sigma)\ln\left[\frac{P_{ijt}}{P_{Ijt}}\right].$$

Substituting the term defined by (3) into β_{ijt} and β_{Ijt} respectively produces

$$(19)\quad \ln\left[\frac{f_{ijt}}{f_{Ijt}}\right] = (\sigma-1)\ln\left[\frac{\prod_k Q_{ikt}^{\gamma_k Z_{jt}} \prod_h G_{ijht}^{\gamma_h Z_{jt}}}{\prod_k Q_{Ikt}^{\gamma_k Z_{jt}} \prod_h G_{Ijht}^{\gamma_h Z_{jt}}}\right] + (1-\sigma)\ln\left[\frac{P_{ijt}}{P_{Ijt}}\right].$$

Equation (19) simplifies to

$$(20)\quad \ln\left[\frac{f_{ijt}}{f_{Ijt}}\right] = (\sigma-1)\left[\sum_k \gamma_k Z_{jt}\ln\left[\frac{Q_{ikt}}{Q_{Ikt}}\right] + \sum_h \gamma_h Z_{jt}\ln\left[\frac{G_{ijht}}{G_{Ijht}}\right]\right] + (1-\sigma)\ln\left[\frac{P_{ijt}}{P_{Ijt}}\right].$$

which is linear in the parameters γ_k and γ_h. In (20) the restriction $-1 \leq \gamma_k Z_{jt} \leq 1$ over the potential range of the quality variables Q_{ikt} for each k assures diminishing marginal benefits from any quality increase at any site.

Equation (20) was estimated by weighted least squares regression. Weights were the square root of the sample size of respondents from each zone of origin. Large sample size is weighted more heavily to reflect a smaller variance in number of sampled visits. With 18 site quality variables, 7 demographic variables, and 1 elasticity of substitution among pairs of sites, 26 parameters are estimated in total.

RESULTS

Table 1 shows the results of parameter estimates in (2) - (4). As expected price effects are strong both in magnitude and statistical significance. The elasticity of substitution characterizes the effects on a zone's visitation patterns among sites, which is a result of the distribution of their travel costs. Its estimated value is 1.63, with a t-statistic greater than 18. That is, increasing an entry fee at a given site has a strong percentage effect on redistribution visitation among sites.

The estimated elasticity of substitution is high and statistically significant because it is the only parameter that appears in the model that accounts for both own and substitute site price effects. All price effects in the model focus on this single parameter. This focus is both a strength and a limitation of the chosen model specification. It is a strength because this specification allows an arbitrarily high number of sites in the model without having to bring in new price parameters for each added site. One can justify this identical price structure across sites by thinking of each site is an identical good, differentiated only by its price and characteristics. The identical price structure is a limitation because a considerable amount of price structure is forced on the model without permitting different sites to reveal unique own- and cross-price effects.

All site quality coefficients shown in Table 1 also have the expected sign. Improving quality characteristics at a site generally seen as more attractive to anglers increases their visits to the site and draws visits away from substitutes. Among the fishing quality variables both the average number of fish caught per day and the average weight of fish kept are marginally significant at best. Negative factors, such as precipitation, water turbidity, and macrophytes reduce visits at a given site and redistribute them to substitute sites. On-site temperature entered with a negative coefficient. This result occurs because most fishing in New Mexico occurs in the spring and summer, when desirable cooler temperatures generally occur at the higher elevations.

Surface area of the water body entered the model strong, both in sign and statistical precision. The parameter estimate of size in water was 1.42, with a t-statistic of more than 10. The size of each water body was measured in surface acres, based on the principle that greater surface area tends to produce greater shoreline access and greater space for boat anglers. By choosing surface area as the water quantity variable, both streams and lakes could be accommodated in the same data set.

DISCUSSION

Economically efficient natural resource policy requires the comprehensive management of diverse opportunities over a complete system of sites. Restricting use and benefits analysis to individual isolated program elements produces the wrong mix of program elements. The result is that complementary elements are underfunded and substitute elements are overfunded.

Effective comprehensive planning simultaneously considers management proposals at all sites, all quality dimensions, and all time periods. Comprehensive management also requires more cooperation among agencies. In New Mexico, at least five federal agencies and two state agencies control recreational opportunities that can alter the estimated benefits of angler-oriented management. Multi-site models, such as the one presented in this paper, could reduce the number and cost of program contradictions if it were accessible to all agencies in a cooperative environment.

Table 3 shows that the incremental benefits of an improvement at any given site depends on conditions at all sites. These results illustrate how program contradictions can be controlled through a demand systems approach to valuing natural resource improvements. Marginal benefits of program elements in Table 3 are sensitive to the context in which their improvements occur. Existing policies that affect conditions at all sites influence the value of a policy that alters conditions at a given site.

Under baseline conditions that occurred in 1992 (Table 3 column 2) marginal values per added acre foot of water used to augment streamflow for sport fishing vary widely. Values in column (3) range from a low of $4 at the Upper Rio Grande to a high of $101 at the Upper Pecos River. Pecos River marginal values are higher because the site has few nearby substitutes, relatively low flows, high fishing quality to complement the streamflow, and it is close to the major population corridors of Santa Fe and Albuquerque. The upper Rio Grande is more isolated, has considerably more base streamflow, more nearby substitutes (such as the Rio Chama) and lower quality fishing.

Marginal benefits under column (4) indicate incremental values of added streamflow when flows are reduced by half of 1992 levels at all sites, while fish catch and other facilities are maintained at 1992 levels. Part of the increase in marginal value compared to the baseline values in column (3) is due to lower flows at a given site, and the remainder is due to reduced flows at all the substitute sites. For example, the $25 per acre foot marginal value at the Upper Rio Chama would be much lower if flows at the other 10 substitute sites were held at 1992 levels. Holding flows constant at a given fishing site is worth considerably more under conditions when policies or nature have reduced flows at important substitute sites.

Column (5) indicates values per added acre foot, if, in addition to the scenario under column (3), fish catch rates are doubled to compensate for low flows. As expected, because fishing quality at a given site complements streamflow, improving fish quality increases the marginal fishing value of added streamflow.

Marginal benefits under columns (6) - (7) have an analogous interpretation as (4) and (5), but flows are doubled with constant fishing quality and with reduced quality fishing, respectively. As expected, the value of added streamflow at any site is much lower in wet than in dry periods. Part of the reduction in streamflow values at a given site, such as the Upper Chama, occurs because substitute sites are also wet. The $7 value of added flow at the Chama would be considerably higher if management actions at other sites produced average or low flow.

Results in Table 3 illustrate the importance of cooperation among agencies. Agencies responsible for controlling streamflows are typically water managers, such as the Forest Service, Bureau of Reclamation, or Corps of Engineers, while other agencies, such as State Game and Fish or Parks and Recreation departments, manage fishing and other site qualities. State departments could better allocate resources for improvements within and across sites if they were aware of federal plans for policies, such as timber harvest schedules, that affect streamflows. Values of added flow for fishing depend strongly on which site has the improved flow, general water supply conditions, and management plans at all sites. The optimal plan over time and space for controlling reductions in streamflows depends on fish stocking plans over time and space.

This paper has opened a door, but much remains to be done. Probably the greatest limitation of this paper is that the specified demand system is only partial. There is no substitution among the total consumer budget and expenditure on fishing resulting from nature, management, or the rest of the economy. Because consumers cannot substitute out of other goods and into fishing expenditures, the benefits of fishing improvements are understated (Hanemann and Morey, 1991). A complete demand system would also account for the effects on marginal values of improvements resulting from outside economic changes, such as increases in gasoline prices or increased leisure time. More work needs to be done that allows demand systems with quality to test for sets of complements versus substitute elements. The model in this paper assumes that all sites are substitutes for each other, while all quality elements at a given site are complements. While this is likely to be true in most cases, one can imagine exceptions. Nevertheless, we believe that conducting cost-benefit analysis based on utility theoretic demand systems are an important step towards reducing program contradictions.

REFERENCES

Cole, Richard A. and Frank A. Ward. "Sustaining Reservoir Fisheries and Angler Benefits Through Comprehensive Management Analysis," *American Fisheries Society*, Forthcoming (1995).

Cole, Richard A. and Frank A. Ward. "Optimum Fisheries Management Policy: Angler Opportunity versus Angler Benefit," *North American Journal of Fishery Management*, February (1994) 14:22-33.

Hanemann, W.M. and E. Morey. "Analysis of Partial Systems of Demand Equations," *Journal of Environmental Economics and Management*, March (1991):20, ------.

Hoehn, John P. and Alan Randall. "Too Many Proposals Pass the Benefit Cost Test," *American Economic Review*, June (1989) 79:544-551.

Lave, Lester B. "Controlling Contradictions Among Regulations," *American Economic Review*, June (1984) 74:471-76.

Lichtenberg, Erik and David Zilberman. "The Welfare Economics of Price Supports in U.S. Agriculture," *American Economic Review*, December (1986) 76:1135-41.

Pollack, Robert A and Terence J. Wales. *Demand System Specification and Estimation*, New York, Oxford University Press, 1992.

Table 1. Site Facility Parameter Estimates, New Mexico Fishing Waters. Estimates are obtained from OLS Estimation of Log Transformed CES Demand System,

$$ln \frac{[f_{ijt}]}{[f_{Ijt}]} = (\sigma-1)\ \{\sum_k \gamma_K\, Z_{jk}\ \ln\frac{[Q_{ikt}]}{[Q_{Ikt}]} + \sum_k \gamma_k Z_{jk}\ ln\frac{[G_{ijkt}]}{[G_{Ijkt}]}\} + (1-\sigma)\ \ln\frac{[P_{ijt}]}{[P_{Ijt}]}$$

where f_{ijt}/f_{Ijt} is the ratio of expenditures on the ith site jth zone of origin tth season to expenditure at the reference site jth zone and tth season. The Ith reference site, Elephant Butte Reservoir, is located in south central New Mexico.

Variable no. (k)	Quality Indicator (Q_k)	Units	Parameter estimate (γ_k)	Student t ratio
1	Access	Percent of shoreline accessible within a 1/4 mile walk from vehicle	0.03	(-)[1]
2	Average weight kept	Average weight of harvested fish (grams); game fish weighted twice pan fish	0.30	(-)
3	Boat Ramps	Number of concrete boat ramps at a site	11.40	(11.81)
4	Developed Campsites	Number of developed campsite within 10 miles of site	0.71	(2.97)
5	Drinking Water	0-1 dummy; 1 = drinking water available at site	1.25	(0.96)
6	Fish/day Kept	Fish number harvested per day; game fish weighted twice pan fish	6.39	(6.73)
7	Fish/day Caught	Fish number caught per day; game fish weighted twice pan fish	0.24	(1.93)
8	Macrophyte	Water plant vegetation, rated by fishery managers from 0 (none) to 10 (highest density)	-0.002	(-)
9	Precipitation	Seasonal precipitation at site, inches	-10.77	(-16.68)
10	Modern Toilet Access	Number of modern toilets within 10 miles of site	3.93	(9.41)
11	Surface Area	Site temperature by season	1.42	(10.53)
12	Tailwater	Water turbidity rated by fishery manages, 0 (none) to 10 (muddy)	0.02	(-)
13	Tempsite		-10.18	(-3.94)
14	Turbidity		-0.002	(-)

[1] Student t-ratios not defined, parameter entered as a restriction.

Table 1. Continued. Site Facility Parameter Estimates, New Mexico Fishing Waters. Estimates are obtained from OLS Estimation of Log Transformed CES Demand System,

$$ln \frac{[f_{ijt}]}{[f_{Ijt}]} = (\sigma-1) \{\sum_k \gamma_K Z_{jk} \ln\frac{[Q_{ikt}]}{[Q_{Ikt}]} + \sum_k \gamma_k Z_{jk} \, ln \frac{[G_{ijkt}]}{[G_{Ijkt}]}\} + (1-\sigma) \ln \frac{[P_{ijt}]}{[P_{Ijt}]}$$

where f_{ijt}/f_{Ijt} is the ratio of expenditures on the ith site jth zone of origin tth season to expenditure at the reference site jth zone and tth season. The Ith reference site, Elephant Butte Reservoir, is located in south central New Mexico.

Variable no. (h)	Quality Indicator (G_{ijht})	Units	Parameter estimate (γ_h)	Student t ratio
1	Cold River	Large cold stream fishing site	6.16	(0.93)
2	Forest Lake	Nearest lake within the boundaries of a National Forest to a city with population of 50,000	27.70	(3.47)
3	Kokanee Lake	Lakes that support Kokanee Salmon with area less than 10,000 surface acres	22.89	(3.56)
4	Mid-Elevation Lake	Lakes with elevation between 4000 feet and 7000 feet and have 1 or 2 boat ramps	26.67	(3.44)
-	-	elasticity of substitution (σ)	1.63	18.91

[1] Student t-ratios not defined, when parameter entered as a restriction. All t-ratios based on log-ratios of actual quality to reference quality.

Table 2. Zone Index, Z_{jt}, as a Multiplicative Function of Individual Demographic Variables, New Mexico Fishing Model.

$$Z_{jt} = \prod_{L} (D_{Ljt})^{\alpha_L} .$$

Variable number (L)	Demographic Variable (D_{Ljt})	Units	Parameter estimate (α_L)	Student t ratio	Approx. p > $\lvert t \rvert$
1	HISP	Percent of persons by zone of Hispanic origin.	-0.42	(-6.15)	0.0001
2	AGE65	Percent of persons by zone over 65 years of age.	1.67	(6.47)	0.0001
3	COLL	Percent of persons by zone under 25 years of age with college education.	1.07	(16.34)	0.0001
4	SFWC	Percent of households by zone headed by single female with children.	-1.30	(-3.01)	0.0026
5	MCNC	Percent of households by zone of married couples without children.	-2.17	(-4.05)	0.0001
6	MCWC	Percent of households by zone of married couples with children.	3.31	(5.57)	0.0001
7	SMNC	Percent of households by zone of single male without children.	0.54	(5.11)	0.0001

Table 3. Instream flow marginal values of water for selected New Mexico for fishing streams, under varying conditions.

(1) Stream	(2) 1992 Ave. Summer Surface acres	(3) Baseline[1]	(4) Dry[2]	(5) Dry[3]	(6) Wet[4]	(7) Wet[5]
		1990 dollars per added acre foot				
Upper Rio Chama	3079	11	25	31	7	4
Middle Rio Chama	254	7	14	29	2	3
Lower Rio Chama	336	16	34	35	8	7
Gila River North Fork	326	14	26	26	6	5
Gila River, South Fork	513	7	14	14	3	3
Gila River, West Fork	57	58	82	83	21	18
Upper Pecos River	174	101	199	201	60	48
Upper Rio Grande	1369	4	7	10	1	1
Middle Rio Grande	828	13	25	25	6	6
San Juan River, East	608	72	135	138	37	35
San Juan River, West	1439	6	12	12	3	3

1 = Streamflow, fish catch rates, and site facilities at 1992 levels
2 = Change all streamflows to 0.5 1992 levels, fish catch and facilities = 1992 levels
3 = Change all streamflows to 0.5 1992 level, double 1992 fish catch, facilities = 1992 levels
4 = Change all streamflows to 2.0 1992 level, fish catch and facilities = 1992 levels
5 = Change all streamflows to 2.0 1992 level, halve 1992 fish catch, facilities = 1992 levels

An Estimate of the Economic Value of Selected Columbia and Snake River

Anadromous Fisheries , 1938-1993

by

Edgar L. Michalson

Introduction

The goal of this paper is to estimate the economic value of the Columbia River anadromous fisheries in the Snake, Clearwater, Grand Ronde, and Salmon rivers. The first dam on the Columbia River was built in 1939 and the last dam was completed in 1968. The first dam on the Snake River was completed in 1962, and the last one in 1975. Since 1938 anadromous fish have had to navigate past as many as eight dams as more dams were constructed. The U.S. Army Crops of Engineers has maintained fish ladders and counting facilities, and has provided an annual count of the numbers of fish by major species since 1938[1]. The species of fish counted and included in this study include chinook and sockeye salmon, and steelhead trout. Over the years the number of fish counted annually has varied greatly. This variation is related to factors such as the cycle of spawning, river and ocean conditions, and a number of other factors which are not well understood. Among these factors is the nature of the particular species to respond to changing environmental conditions which is reflected in that some mature fish return to spawn after one year (jacks), some after two years, some after 3 years, some after four years, and some after 5 years. This variability is reflected in the fish counts and appears as the peaks and valleys on the charts which follow. This phenomenon applies to all of the fish runs on the Columbia River System. The U.S. Army Corps of Engineers has contracted with the Washington Department of Wildlife to do fish counting. Under this

[1] U.S. Army Corps of Engineer District. "1990 Annual Fish Passage Report: Columbia and Snake Rivers for Salmon, Steelhead, and Shad." North Pacific Division Corps of Engineers, 1990.

contract fish are counted according to a preset schedule based on one or two eight hour shifts. Fish were counted for 50 minutes, and these totals were multiplied by a factor of 1.2 to estimate the passage for the full hour. Each dam on the system has a specific schedule which is based on observed passage information which has been developed over the span of time involved for each dam.

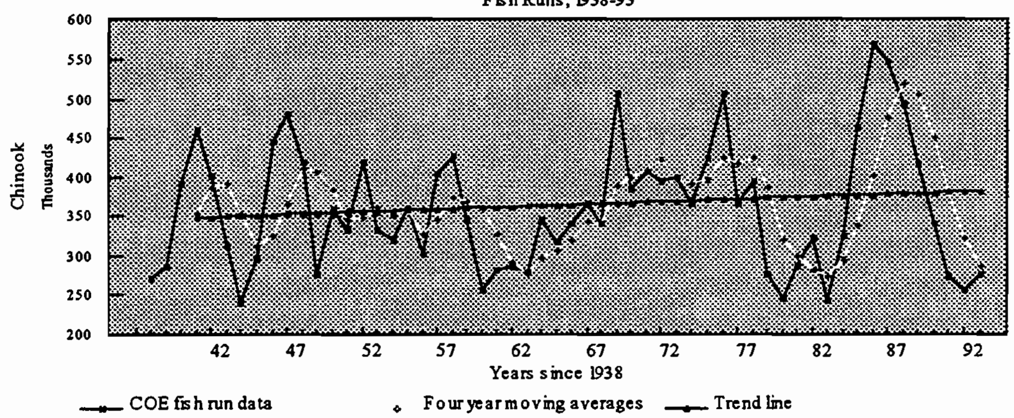
Fish Passage

Since 1938 fish passage for anadromous fish have been charted for each dam. A simple linear trend line has been computed for each species and dam from the year the reservoir was filled until 1993. The chinook and steelhead fish runs passing the Columbia River dams have showed an upward trend since 1938. The fish runs for sockeye salmon show a downward trend over this same time period. At the Snake River dams the chinook and sockeye salmon both show a strong downward trends, while the steelhead trout show a strong upward trend. The linear regression equations used in this study are not entirely satisfactory in that the regression coefficients are all very close to zero which tells us that when time is used as an explanatory variable it doesn't explain very much of the variation in fish numbers. This is also evident in the great variability shown in the annual fish passage numbers. The peaks and valleys of the graphs presented in figures 1 though 9 tend to increase as one moves upstream. And, on the Snake River the variability is much greater than it is on the Columbia River. This variability is in some cases 6 or 7 times that of the lowest fish passage numbers reported. A great deal of this variability is also related to the spawning phenomenon discussed above. One of the obvious reasons that there has been a positive trend related to steelhead undoubtedly has been the successful operation of the steelhead hatcheries in Idaho, Oregon, and eastern Washington which have supplemented the wild steelhead runs.

The run pattern for chinook salmon on the Columbia River over Bonneville Dam shows an increasing trend over the 55 year period that fish have been counted. Figure 1 shows both the Corps of Engineers annual fish run data, the four year moving average, and the trend for the chinook fish runs since 1938. The year to year variability is great as evidenced by the peaks and valleys over the 55 year time period. The annual fish runs seem to have varied greatly over time with the amplitude of the peaks and valley's ranging from 240,000 to a high of 571,000 fish

Figure 1. Bonneville Dam Chinook

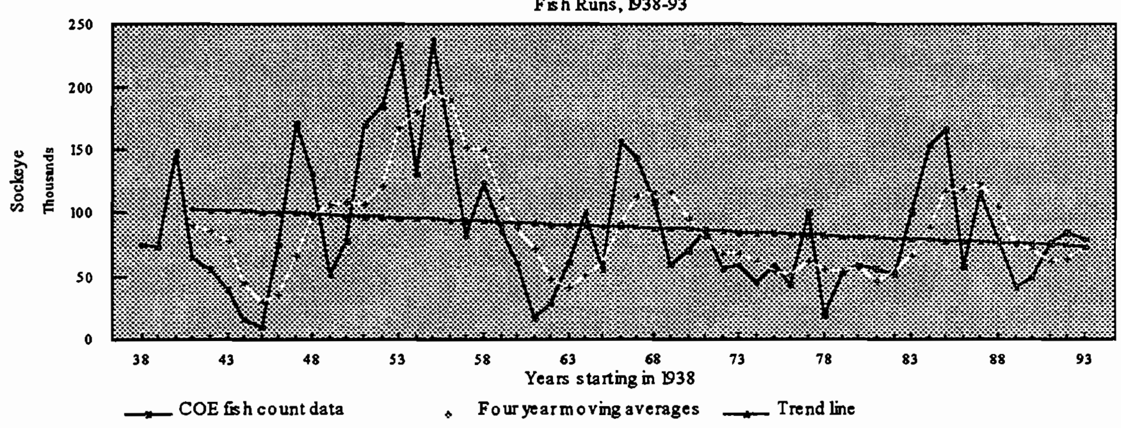
Fish Runs, 1938-93



The run pattern for sockeye salmon on the Columbia River show a declining pattern since 1938, see figure 2. The decline as measured at Bonneville Dam is approximately 20 percent over the 55 year period. The variability in sockeye fish runs ranges from a low of 9,501 in 1945 to a high of 237,748

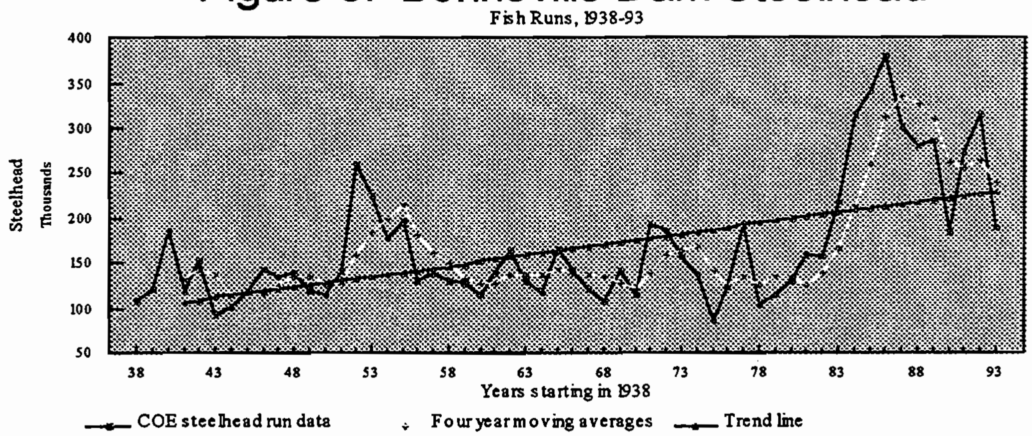
Figure 2. Bonneville Dam Sockeye

Fish Runs, 1938-93



in 1955 and 235,215 1953. Since the mid 1950's these runs have been on a long term decline. There have been minor highs and lows since then, but the range has been narrowed, and none of the recovery periods have exceeded 170,000 fish on an annual basis.

Figure 3. Bonneville Dam Steelhead

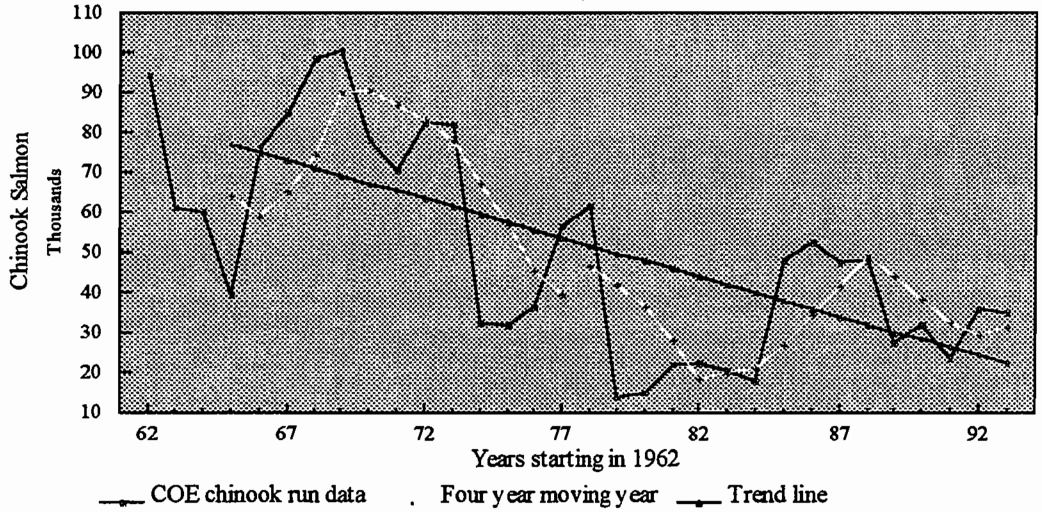


The annual tally for steelhead begins at slightly over 100,000 fish in 1938 at Bonneville and ends up at slightly over 200,000 fish in 1993. The variability in annual fish count numbers ranged from a low of approximately 80,000 fish in 1975 to a high of almost 380,000 fish in 1985. There appears to have been along term decline from 1938 to 1978, and a significant recovery after 1978 at Bonneville Dam. This recovery seems to correlate reasonably well with the time that the Ahsaka Fish Hatchery began producing steelhead trout..

The anadromous fish runs on the Snake River tend to increase at a slower rate than those on the Columbia River based on the trend estimates. At Ice Harbor Dam, the trend estimate was 76,899 for chinook salmon in 1962 and decreased to 22,355 fish in 1993. There has been approximately a 50 percent decline in chinook salmon runs over Ice Harbor Dam as estimated between 1962 and 1993.

Figure 4. Ice Harbor Chinook

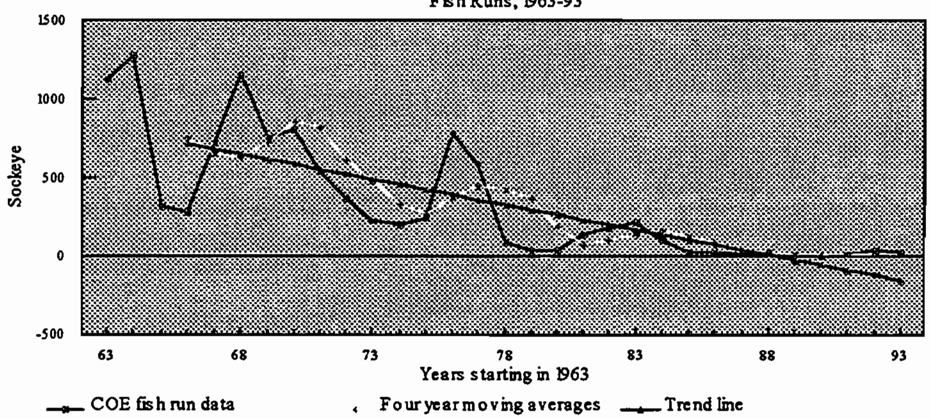
Fish Runs, 1962-93



The last Sockeye counted over Ice Harbor Dam was counted in 1989, and this was also true for the other upriver dams. The sockeye salmon appears to have disappeared from the Snake River, and may truly be an endangered species.

Figure 5. Ice Harbor Dam Sockeye

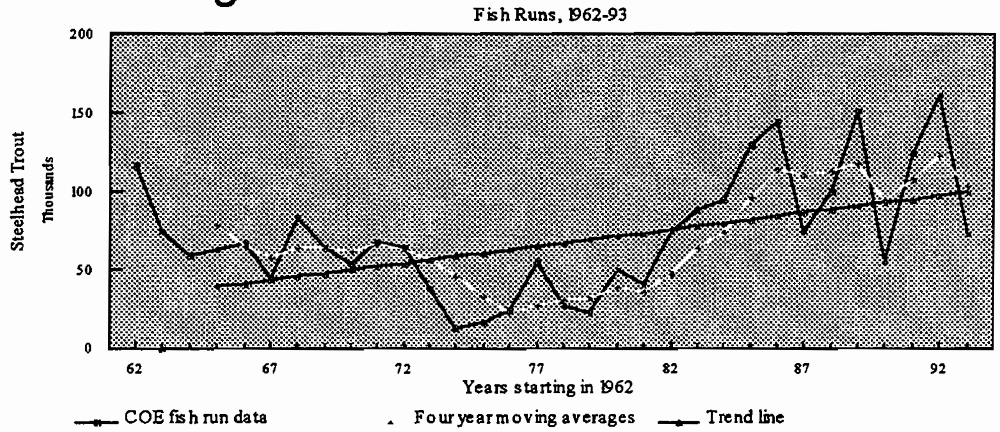
Fish Runs, 1963-93



The graph of fish run data speaks clearly as to what has happened to the sockeye salmon. The trend line analysis indicated that there were only 710 fish returning to spawn and since 1988 the expectation is that none will return. The same pattern set for sockeye salmon was followed at Lower Granite Dam (1975-1993) over the period of time that it has been in operation.

The steelhead runs over Ice Harbor Dam have tended to increase over the period from 1962 until 1993 as show in figure 6. The trend line based on the four year moving averages begins with 39,131 fish

Figure 6. Ice Harbor Dam Steelhead

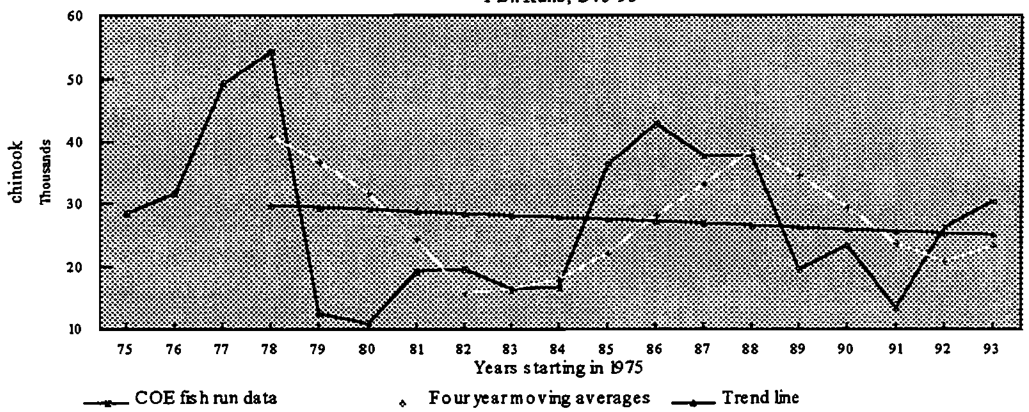


in 1962 and increases to 98,939 fish in 1993. The actual fish counts declined from 1962 to a low of 12,528 fish in 1974, but since 1974 the number of fish returning to spawn has been on an increasing trend. The maximum number of steelhead counted over Ice Harbor Dam occurred in 1992 with over 160,000 fish passing the dam.

The pattern of chinook salmon counts at Lower Granite Dam follow those at Ice Harbor Dam, however, the trend in fish numbers is down, but at a much lower rate than at Ice Harbor. The trend estimate was 29,749 chinook in 1978, and had declined to 24,964 in 1993. The downward slope of the trend line indicates that there has been a problem

Figure 7. Lower Granite Dam Chinook

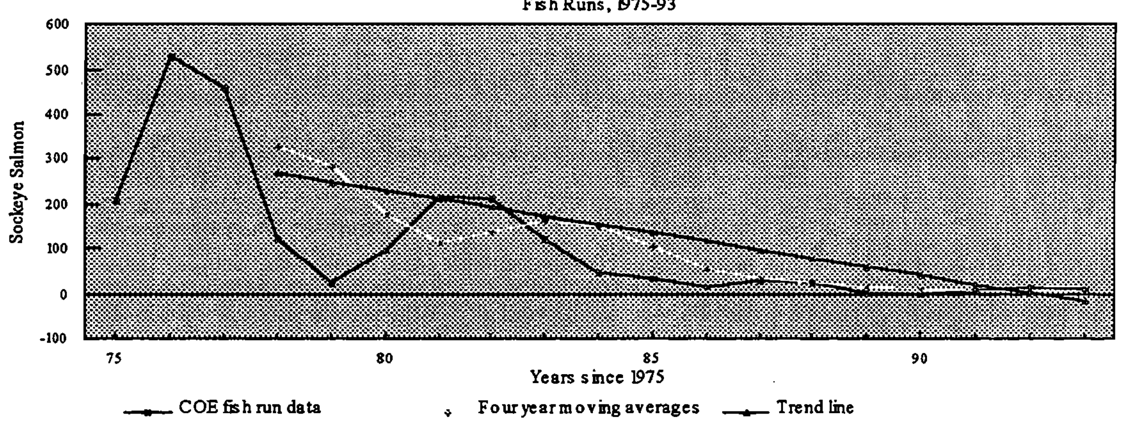
Fish Runs, 1975-93



with fish passage between Lower Granite Dam and Ice Harbor Dam. In addition to the Lower Granite Dam the chinook salmon also have to pass two other dams down stream on the lower Snake River. These are Lower Monumental Dam and Little Goose Dam. The chinook salmon (and all other anadromous fish passing Lower Granite Dam) have to pass eight dams total to get to their spawning reds (or to the fish hatcheries).

Figure 8. Lower Granite Dam Sockeye

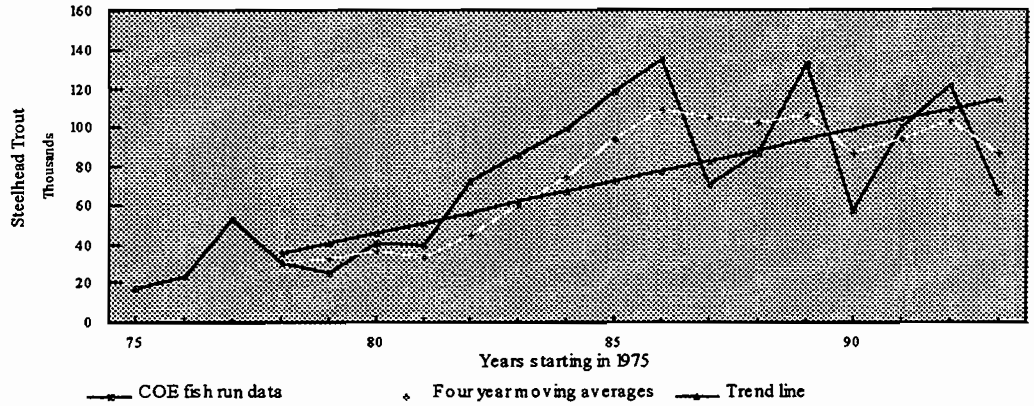
Fish Runs, 1975-93



The sockeye salmon are extinct in Idaho according to this analysis. According to the trend line estimate, there were only 269 fish passing Lower Granite Dam in 1978, and this compares to 326 fish passing Ice Harbor Dam. The loss due to fish passage amounts to over 17 percent or approximately 8 percent at each dam between Ice Harbor Dam and the Lower Granite Dam.

Figure 9. Lower Granite Dam Steelhead

Fish Runs, 1975-93



The case for the Lower Granite Steelhead is very different that that for either the chinook or sockeye salmon. The steelhead trout are obviously able to adapt to the changed river and fish passage conditions much better than the other species. The trend estimate in 1978 was 35,416 fish and in 1993 that had increased to 114,736 fish. This amounts to a 324 percent increase in the number of steelhead over a 15 year period. A rather impressive accomplishment. If one looks at the dispersion of data around the trend line it is also obvious that the over all trend has been upward over this time period. This time period also corresponds to the time when the Ashaka Fish Hatchery was brought into full production. In the case of steelhead it is evident that the hatchery process works very successfully.

The pattern of fish migration on the Columbia and Snake rivers needs to be interpreted in light of the many changes which have occurred over the 55 years or more that the river system has been operated. These changes include the development of irrigation, the addition of dams to the river system, population growth in the region, increased timber harvest levels, and the expansion of the river as a recreational resource. It also reflects an increasing reliance on fish hatcheries to supply fish to the river system to make up for the short fall of natural fish production. This is especially clear in the case of the chinook salmon and the steelhead trout. The development of fish hatcheries for these species has contributed more than half of the fish counted, and does indicate a severe decline in native wild fish stocks for these species.. . The slope of the trend lines for steelhead trout passing the upstream dams tends to reflect the impact of hatcheries, spills, and barging. It is interesting that the chinook returning to Idaho have declined less than 10 percent between 1978 and 1993 as measured by the trend line as shown in figure 9. This should be compared with the more than 70 percent decline in chinook salmon counted at Ice Harbor dam. It the trend estimates for 1978 at both dams are compared for chinook salmon, the estimate was 51,575 at Ice Harbor Dam, and 29,749 at Lower Granite Dam. It is clear that most of the fish lost on the Snake River have been lost between Ice Harbor Dam and Lower Granite Dam. The major losses seem to have occurred on the Tucannon River in eastern Washington.

Estimated Value of Columbia and Snake River Fisheries

The question arises as to the worth of these fish which are part of the natural environment, and which existed prior to the building of the dams on these rivers. This question has become even more pressing with the threat of naming some of these fish endangered species. Arguments are being made that the economic losses from the strategies being proposed far outweigh the value of these fish, but little evidence is presented on the side of the fish. The approach will be to estimate the value of the salmon and steelhead fisheries in the Snake River system. The approach used to estimate fishery value was based on the American Fisheries Society document entitled "Monetary Values of Freshwater Fish and Fish-Kill Counting Guidelines published in 1982. The "Monetary Values of Freshwater Fish Committee" and the "Pollution Committee" set forth procedures and estimated the value of each specie of freshwater fish.[2] The values developed in the above study have been updated from 1982 to 1993 using the producer price index. [3] The Value per fish are shown in Table 1 along with the average weight, the estimated

value per fish, and the capitalized value per fish. This latter value may be interpreted as that amount of money which would have to be set aside to generate the corresponding

[2] The Monetary Values of Freshwater Fish Committee and the Pollution Committee. "Monetary Values of Freshwater Fish and Fish-Kill Counting Guidelines." American Fisheries Society, Special Publication No. 13. 1982.
[3] USDA. "Agricultural Outlook." March 1992/AO-183.

cash flow generated by the fish value. A four percent "social discount rate" was used to determine the capitalized value of these fish. Recognizing that the choice of discount rate

Table 1. Monetary Values of Colombia and Snake River Anadromous Fish, 1993.

Species	Weight lb.	Value* per lb. $	Value per fish $	Capitalized value $
Chinook Salmon	33	$.4.96	$163.68	$4,092
Sockeye Salmon	10	$4.96	$49.60	$1,240
Steelhead trout	17	$4.96	$84.32	$2.108

Source: American Fisheries Society. "Monetary Values of Freshwater Fish and Fish-Kill Counting Guidelines." American Fisheries Society . Special Publication No. 13, 1982.

Adjusted by the Producer Price Index to update the values from 1982 to 1993.

is arbitrary, and that one could argue for higher or lower discount rates, this discount rates should not be considered to be set in concrete, but rather as a first step in the process of determining the value of these fish.

The average replacement cost of each species is shown in table 1. Each chinook salmon was valued at $163.68, sockeye salmon at $49.60, and steelhead trout at $84.32. The capitalized value of each fish for spawning using the 4 percent discount rate was: 1) chinook salmon $4,092, 2) sockeye salmon $1,240, and 3) steelhead trout $2,108.

Columbia River Fisheries

The next step was to calculate the value of the fishery based on the numbers of fish shown in the Corps of Engineers fish count data. These values are shown in table 2 for Bonneville Dam. The values shown are based on the trend analysis for each species of fish over the time period covered by the fish counting procedure. The beginning and the ending values are shown in table 2 and the differences are were computed. At Bonneville Dam the trend for chinook salmon and steelhead trout was increasing, and that for the sockeye salmon was decreasing. The changes in the fisheries over time have tended to be increasing in the aggregate. The total number of fish has increased from an estimated 567,865 fish in 1941 to 685,741 fish in 1993. This represents an overall increase of over 114,000 fish, or a 20% increase over 51 years. The economic values were computed using the values in table 1. The annual value of all of the fisheries at Bonneville Dam was estimated to have increased between 1941 and 1993 by $14.3 millions. The total stock value in 1941 was $71.1 millions and it had increased to $85.4 millions by 1993. The stock value (or capitalized value) of these fisheries had increased by $230 millions. The total stock value in 1941 was estimated at $1.8 billions, and this increased to $2 billions by 1993.

Table 2. Estimated Economic Value of Selected Columbia River Fisheries, 1941-93.

Dam / species	Year	The number of fish	Average Value per fish	The annual value of fishery	Capitalized value of fishery*
			$	mil. $	mil. $
Bonneville					
Chinook	1941	348,133	$163.68	$57.0	$1,425.0
	1993	381,881	(same)	$62.5	$1,562.5
	Difference	33,748	NA	$5.5	$137.5
Sockeye	1941	103,126	$46.90	$5.1	$127.5
	1993	74,058	(same)	$3.5	$87.5
	Difference	(29,068)	NA	($1.6)	($40.0)
Steelhead	1941	106,562	$84.32	$9.0	$225.0
	1993	229,802	(same)	$19.4	$485.0
	Difference	123,240	NA	$10.4	$260.0
Totals	1941	567,876	NA	$71.1	$1,777.5
	1993	685,741	NA	$85.4	$2,007.5
	Difference	114,865	NA	$14.3	$230.0

- **A four percent discount rate was used in this study.**

In terms of the individual species between 1941 and 1993: 1) the annual value of the chinook salmon increased from $57 millions to $62.5 millions, 2) the annual value of sockeye salmon declined from $5.1 millions to $3.5 millions, and 3) the annual value of steelhead trout increased from $9.0 millions to $19.4 millions. The net gains were: 1) chinook salmon, $ 5.5 millions, 2) sockeye salmon, a negative $1.6 millions, and 3) steelhead trout, $10.4 millions. The stock value of the individual fisheries also changed between 1941 and 1993 as follows. Chinook salmon increased by $137.5 millions from

$1.425 billions in 1941 to $1.562 billions in 1993. The sockeye salmon's stock value declined by $40 millions from $127.5 millions in 1941 to $87.5 millions in 1993. The steelhead trout which showed the largest gain increased by $260 millions from $225 millions to $485 millions in 1993. The general conclusion is that the aggregate value of these fisheries has increased slowly since Bonneville Dam was completed. It is also true that most of this increase has resulted from the mitigation in terms of hatcheries, barging, and fish passage improvements that have been added to the systems since 1938.

Lower Snake River Fisheries

The next point at which values were measured was at Ice Harbor Dam near the mouth of the Snake River. Idaho, Oregon, and Washington up river chinook salmon, sockeye salmon, and steelhead trout all pass through this dam. The major contributors to this value were the chinook salmon and steelhead trout. The major loss in value was that related to the decline in the number of chinook salmon in the river. The total number of chinook declined by 54,534 fish over this time period, or a loss of 70 percent. The number of steelhead trout has increased by almost 60,000 fish in the same time period, or an increase of 153 percent. The sockeye salmon the loss was not economically significant, but was environmentally a disaster for this fishery because for all purposes the run has been lost..

The value of the Snake River fishery is shown in table 3. The estimated value of the Snake River fisheries in 1965 was $397.8 millions in terms of its stock value, and it generated an annual flow worth $16.3 millions. In 1993 the estimated stock value of these fisheries had declined $297.5 millions, a loss of $100.3 millions, and the annual flow to $11.9 millions. In terms of the individual fisheries the greatest decline occurred to the chinook salmon which declined in terms of its estimated stock value from $315 millions to $90 millions. The annual value of the chinook salmon fishery declined $225 millions in terms of its stock value, and $9 millions in terms of its annual value.

The sockeye salmon on the Snake River never was a large fishery, and it declined to zero by 1988. There has been a complete loss of this fishery which had a stock value in 1962 of $0.8 millions and annual value of $0.4 millions.

Table 3. Estimated value of Selected Lower Snake River Fisheries, 1965-93.

Dam /species	Year	Average number of fish	Value per fish	The annual value of fishery	Capitalized value of fishery*
			$	Mil $	Mil. $
Ice Harbor					
Chinook	1962	76,889	$163.68	$12.6	$315.0
	1993	22,355	(same)	$3.6	$90.0
	Difference	(54,534)	NA	($9.0)	($225.0)
Sockeye	1963	710	$49.60	$0.4	$0.8
	1993	(0)	(same)	$0.0	$0.0
	Difference	(710)	NA	($0.4)	($0.8)
Steelhead	1962	39,131	$84.32	$3.3	$82.0
	1993	98,939	(same)	$8.3	$207.5
	Difference	59,808	NA	$5.0	$125.5
Totals	1962-3	116,730	NA	$16.3	$397.8
	1993	121,294	NA	$11.9	$297.5
	Difference	4,564	NA	$4.4	($100.3)

*** A four percent discount rate was used in this study.**

The steelhead trout on the other hand are a very different picture. The numbers The steelhead trout on have increased dramatically on the Snake River since 1962. The estimated stock value of these fish has increased from $82 millions in 1962 to $207.5 millions in 1993, and the estimated annual value from $3.3 millions to $8.3 millions over the same period. The net increase in terms of stock value has increased by $125.5 millions, and the annual value by $5 millions.

In the case of the Lower Snake River the total number of fish over the Ice Harbor Dam increased between 1962/63 and 1993. The reason for this was the large increase of steelhead trout which occurred. The total number of steel trout increased from approximately 39,000 to over 98,000 fish during the 30 year period, a net gain of 59,000 fish. At the same time the number of chinook salmon declined by over 54,000 fish, and the sockeye salmon disappeared from the river. The percentage increases in decreases were: 1) chinook salmon decreased by 29 percent , 2) sockeye 100 percent, and 3) steelhead increased by 250 percent. The lower Snake River fisheries lost value during this time, mainly because the economic value lost by the decline of the chinook fishery was greater than the gain in terms of the number steelhead trout fishery.

Up River Snake River Fisheries

The last dam to be considered is the Lower Granite Dam on the Snake River. Lower Granite dam is the gateway to Idaho and the upriver Oregon and Washington fisheries. The results of this analysis are interesting in that the chinook salmon runs over Lower Granite although declining over time are not nearly as reduced as those over Ice Harbor Dam. The value of the upstream Idaho, Oregon, and Washington fisheries are shown in table 4. In terms of stock value the chinook salmon have declined by about 20 percent, from a high of $122.5 millions in 1978 to $102.5 millions in 1993. The annual value of the chinook fishery has declined from $4.9 millions to $4.1 millions. This should be compared to the loss at Ice Harbor dam which was approximately 70 percent.

Table 4. Estimated Value of Selected Upper Snake River Fisheries, 1978-93.

Dam /species	Year	The number of fish	Average value per fish	The annual value of the fishery	Capitalized value of the fishery*
			$	Mil. $	Mil. $
Lower Granite					
Chinook	1975	29,794	$163.68	$4.9	$122.5
	1993	24,964	(same)	$4.1	$102.5
	Difference	(4,830)	NA	($0.8)	($20.0)
Sockeye	1975	134	$49.60	$0.007	$0.2
	1993	(0)	(same)	$0.0	$0.0
	Difference	(134)	NA	($0.007)	($0.2)
steelhead	1975	35,416	$84.32	$3.0	$75.0
	1993	114,736	(same)	$9.7	$242.5
	Difference	79,320	NA	$6.7	$167.5
Totals	1975	65,344	NA	$7.9	$236.6
	1993	139,700	NA	$13.8	$345.0

* A four percent discount rate was used in this study.

In the case of the sockeye salmon the results are quite similar to those at Ice Harbor Dam. The number of sockeye passing Lower Granite Dam was even smaller that that for Ice Harbor Dam. The last sockeye passing Lower Granite Dam was counted in 1989. The loss in term of stock value was estimated at $0.2 millions, and in terms of annual value approximately $70,000.

The steelhead trout follow the pattern established at Ice Harbor Dam. The number of fish has increased over 300 percent. The net increase in stock value for steelhead was $167.5 millions, from $75 millions in 1978 to $242.5 millions in 1993. The annual value increased from $3.0 millions in 1978 to $9.7 millions in 1993. The main factor behind this increase is the Ashaka steelhead hatchery on the Clearwater River.

The situation for the Upper Snake River fish runs is markedly different that that for the Lower Snake River fish runs. In this case the changes are not as dramatic as they were for the lower Snake River fish runs. The chinook salmon runs are declining, but at a much slower rate than for the Lower Snake River. The loss of chinook between 1978 and 1993 amount to just under 5,000 fish during the 15 year period. The loss of sockeye was the same as that which occurred at Ice Harbor Dam because the final destination of the sockeye would have been in the upper reaches of the Salmon River drainage in central Idaho lakes. However, for the steelhead trout, an even more rapid increase was shown than that at the Ice Harbor Dam. Steelhead trout numbers increased by almost 80,000 fish during this period.

Conclusions and Observations

The purpose of this paper was to attempt to develop a methodology for estimating the value of the Columbia River fisheries. The important conclusions to be drawn are: 1) that the aggregate fish runs on the Columbia River as counted at Bonneville Dam have tended to increase since 1941; 2) the fish hatcheries, spills, and transport systems which have been used to maintain the up river fish populations have worked reasonably well to maintain and enhance steelhead trout populations on the Snake River; 3) that the greatest loss of fish has occurred on the lower Snake River; and 4) that the sockeye salmon have virtually disappeared from the Snake River.

There is only one spawning stream between Ice Harbor Dam and Lower Granite Dam, the Tucannon River in Washington State, and the greatest loss of fish is undoubtedly related to the loss of habitat on this river. From an environmental point of view, it is clear that the loss of the Tucannon fishery has been a serious blow to the lower Snake River fishery. The loss of over 50,000 chinook salmon on their spawning journey between 1962 and 1993 needs further research.

This study concludes that the aggregate value of the Columbia River fisheries has increased since the Bonneville Dam was built. It is also true that most of this increase has resulted from the mitigation in terms of hatcheries, barging, and fish passage improvements that have been added to the systems since 1938. In terms of the individual species between 1941 and 1993: 1) the annual value of the chinook salmon increased from $57 millions to $62.5 millions, 2) the annual value of sockeye salmon declined from $5.1 millions to $3.5 millions, and 3) the annual value of steelhead trout increased from $9.0 millions to $19.4 millions. The net gains were: 1) chinook salmon, $ 5.5 millions, 2) sockeye salmon, a negative $1.6 millions, and 3) steelhead trout, $10.4 millions. The stock value of the individual fisheries also changed between 1941 and 1993 as follows. Chinook salmon increased by $137.5 millions from $1.425 billions in 1941 to $1.562 billions in 1993. The sockeye salmon's stock value declined by $40 millions from $127.5 millions in 1941 to $87.5 millions in 1993. The steelhead trout which showed the largest gain increased by $260 millions from $225 millions to $485 millions in 1993. The aggregate stock value of these fish runs estimated at Bonneville dam in 1993 was just over $2.0 billions, and the annual value generated was estimated to be $85.4 millions.

The situation on the lower Snake River is mixed in terms of the economic value of the fisheries. The chinook and sockeye salmon are in decline, while the steelhead trout are increasing in a significant way. The value of the chinook salmon on the lower Snake river has declined dramatically since 1965. The stock value of this fishery declined $225 millions, and the annual value by $9 millions between 1965 and 1993. In the case of the sockeye salmon the loss of stock value decline $0.4 millions and an annual value loss of $70,000. The steelhead trout runs however were increasing, and the stock value of this resource increased from $82 millions to $207.5 millions. The annual value of the steelhead runs increased from $3.3 millions to $8.3 millions. The aggregate stock value of the lower Snake River fish runs was estimated to be $297.5 millions in 1993, and this was decline of $100.3 millions from the situation in 1965. The annual value of the lower Snake river fisheries was estimated to be $11.9 millions in 1993, down $4.4 millions from 1965.

The upper Snake River fish runs that feed the Idaho, eastern Oregon and Washington fisheries are being maintained in a better form that those of the lower Snake River. The chinook salmon although in decline have not suffered as much as the lower Snake River chinook salmon. The stock value of this resource has declined from $122.5 millions to $102.5 millions since 1978. The annual value declined from $4.9 millions to $4.1 millions. The fate of the sockeye salmon on the upper river is the same as that for the lower Snake River. In the case of steelhead trout, again the resource is increasing. The stock value of the steelhead trout on the upper river has increased from $75 millions in 1978 to $242.5 millions in 1993. The annual value of the fishery increased from $3 millions to $9.7 millions over the same period. The aggregate stock value of the upper Snake River fisheries was estimated to be $345 millions in 1993, which was up $108.4 millions over 1978. The annual value of the fishery was estimated to be $13.8 millions in 1993, which was up $5.9 millions.

The general conclusion is that the aggregate value of these fisheries has increased since Bonneville Dam was completed. It is also true that most of this increase has resulted from the mitigation in terms of hatcheries, barging, and fish passage improvements that have been added to the systems since 1938. However, it is also clear that some individual species have not fared very well during the 55 years since the first dam was built. The sockeye salmon have been in decline on both the Columbia and Snake rivers, and the chinook salmon have been in serious decline on the Snake River. Steelhead trout however have been increasing significantly on both river systems. In conclusion, some of the fisheries on the Columbia and Snake rivers have been impacted negatively by the development of the dams on these rivers. This negative impact has been mitigated by hatcheries, improved downstream fish passage, improved spawning habitat, and downstream barging of chinook salmon and steelhead trout. In the case of the sockeye salmon, the only hope may be to build hatcheries for these fish, if society wants to ensure their survival in the current river system..

Comparing Economic Values and Behaviors

Between Attitudinal Groups in Alaska

Daniel W. McCollum
(Rocky Mountain Forest & Range Expt. Station)

SuzAnne M. Miller
(Alaska Dept. of Fish & Game)

Rick Sinnott
(Alaska Dept. of Fish & Game)

Presented at the Annual Meetings of Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning, Monterey, CA, March 12 - 15, 1995.

Comparing Economic Values and Behaviors Between Attitudinal Groups in Alaska

Introduction

Social psychologists have long been interested in relationships between attitudes, behavioral intentions, and actual behavior; and the predictability of behavior by attitudes. The concepts and relationships are discussed and summarized in the Fishbein-Ajzen model (Fishbein and Ajzen 1975, and discussed by Mitchell and Carson 1989). Mitchell and Carson cite several studies which have tested the degree to which behavior can be predicted from attitudes. The evidence is mixed. Some studies show strong correlation between attitudes and behaviors, so attitudes do a good job of predicting behavior. Other studies fail to show such correlation.

The conventional wisdom in economics can be summarized as "Actions speak louder than words." One of the criticisms levelled at contingent valuation (CV) is that it does not actually measure behavior; it is just words, so critics are skeptical as to how much credence can be placed on the results of CV studies. This study attempts to test the hypothesis: "Actions speak louder than words, but sometimes they say the same thing." In effect, can measures of attitudes and/or behavioral intentions (of which responses to CV questions are one) be used to predict potentially observable behavior?

Procedure

Two 1992 surveys collected information on how Alaskans thought about and used wildlife. One survey was administered to resident hunters (RH). Hunters were randomly sampled from the population of licensed resident Alaskan hunters. The data consist of 2,077 usable responses, which is a response rate of 58.6%. The focus of the RH survey was on consumptive wildlife use in general, and big game or waterfowl hunting in particular. The survey contained questions on attitudes and opinions about wildlife, wildlife use, and wildlife management, expenditures on hunting related equipment and supplies not attributable to any particular trip, and expenditures and net economic value of one specific overnight big game or waterfowl hunting trip. The second survey was administered to resident voters (RV) as a proxy for the general resident population. People were randomly selected from the list of registered Alaska voters in 1990 (which was the most recent general election), stratified by Alaska state legislative district. The data contain 2,370 usable responses; a 57.2% response. The focus was on nonconsumptive wildlife use in general, and "wildlife viewing trips"[1] in particular. The survey contained questions on attitudes and opinions about wildlife, wildlife use, and wildlife management, expenditures on nonconsumptive wildlife related equipment and supplies not attributable to any particular trip, and expenditures and net economic value of one specific wildlife viewing trip.

---

[1]Wildlife was defined in the survey as including "all wild animals, except fish." Wildlife viewing was defined as "a term we use to mean all activities involving wildlife, <u>except hunting and trapping</u>. Activities such as watching, photographing, tracking, painting, or listening to wildlife are included in the term wildlife viewing."

Both surveys were administered by mail in the Spring of 1992 and asked about wildlife use in 1991. For both surveys, an introductory letter was mailed to all selected individuals, followed by the questionnaire packet to all individuals whose introductory letter was not returned by the Post Office as undeliverable. Two followup mailings were sent to nonrespondents. Chi-square analyses were performed using variables which were known about the population to test for biases in the sample. Weights were calculated to correct for statistically significant differences. The weights were not used in this particular analysis because no attempt is made to infer anything about more general populations. Not using the weights does not detract from insights gained from this analysis.

Cluster analysis was used to classify the Alaskan voters according to their responses to 16 statements dealing with attitudes toward wildlife, wildlife use, and wildlife management.[2] Table 1 shows the 16 statements used to define the groups. Briefly, agglomerative hierarchical clustering (in the SPSS statistical software), performed 50 times on random subsamples of five percent of the sample, identified three or four potential clusters. K-means cluster analyses performed on the entire sample produced usable results for both three- and four-cluster groups. Those analyses presented no strong evidence in favor of either three or four clusters. Miller et al. (1994) discuss four distinct cluster groups. Because they are similar, this analysis combines the two middle clusters and focuses on three clusters.

---

[2]Responses were on a five point scale: strongly agree, moderately agree, moderately disagree, strongly disagree, don't know/no opinion.

Following classification of respondents into clusters or "attitudinal groups," a discriminant function was estimated to predict the attitudinal group to which voters would belong. The discriminant function was based on 13 of the 16 statements shown in Table 1 (those that appeared in both the RH and RV surveys, which are denoted by an asterisk and bold faced type in Table 1). That discriminant function placed respondent voters in the correct attitudinal group 83% of the time. (A discriminant function using the responses to all 16 attitudinal statements placed respondent voters in the correct group 91.5% of the time.) The discriminant function was then applied to respondents to the RH survey, and hunters were classified into the same attitudinal groups as voters.

On some of the statements used to differentiate respondents into attitudinal groups the responses of early respondents differed significantly from those of later respondents, possibly indicating a nonresponse bias. To test for potential nonresponse bias, response rates were examined across groups with regard to when the response was received (i.e., after the first, second, or third mailing). No significant differences were found. While that does not ensure the sample is representative of Alaska voters, group attitudes and other attributes can be compared across attitudinal groups.

Miller et al. (1994) provide more detail about both the statistical procedures used to classify groups, and comparisons between groups in terms of attitudes and demographic characteristics. The groups can be summarized, however, as follows: **Group A's** attitudes toward hunting and consumptive uses of wildlife were generally negative; but their attitudes toward meat hunting and sport fishing were slightly positive. Group A was the strongest proponent of mandatory hunter education for all hunters, and was the most favorable toward using Pittman-Robertson funds[3] for nonconsumptive oriented programs and management. It had the highest proportion of females and urban residents, the shortest length of residence in Alaska, and the highest median incomes and education levels. Group A had the highest proportion of members who had never purchased a hunting license or taken an outing on which wildlife viewing was a planned activity--the highest proportion who had participated in neither activity. Group C, at the other extreme, had the greatest affinity for consumptive uses of wildlife and for hunters. Its members possessed the least positive attitudes toward nonconsumptive wildlife uses, especially when they might affect hunting opportunities; and were opposed to using Pittman-Robertson funds for nonconsumptive programs and management. Group C members were the strongest proponents of bear baiting[4] and most opposed to mandatory hunter education for experienced hunters. Group C contained the lowest proportion of females and its members had the longest residency in Alaska. Group C median income and education levels were similar to Group B, but lower than Group A. Group C had the highest proportion of hunters, the lowest proportion of individuals who had taken a trip on which wildlife viewing was a planned activity, and the second highest proportion with no hunting or wildlife viewing experience--the second highest proportion who had participated in neither activity (but that proportion was not statistically different than that of Group A). Groups A and C had the greatest proportions of nonparticipants in

---

[3]Pittman-Robertson funds are taxes on sales of guns, ammunition, and related supplies collected by the Federal government and allocated to states. A large proportion, but not all, of those taxes are paid by hunters so they are commonly viewed as a revenue source, to state wildlife agencies, derived from consumptive users of wildlife.

[4]The bear baiting question said: "Some people think baiting or attracting black bears with food allows hunters to be more selective in choosing which bear to kill. Do you support allowing hunters to use bait to hunt black bears?"

wildlife-related activities, but the compositions of the participants were different--Group A had more nonconsumptive users with relatively fewer hunters and Group C had more hunters with fewer nonconsumptive users. Group A was the least consumptive oriented group, but there were some hunters in Group A. Group C was the most consumptive oriented group, but not all Group C members were hunters. Group B was between those two extremes in all respects.

Following classification into attitudinal groups, survey information for both hunters and voters related to wildlife participation and wildlife-related expenditures was analyzed by group. Information was tested for statistically significant differences between groups using nonparametric tests. The Mann-Whitney test was used to test differences between the two extreme groups (A and C) and the Kruskall-Wallis test was used to test differences between all three groups.

Results

Table 2 shows the relative proportions of the voter and resident hunter samples falling into each attitudinal group. Both samples show about the same proportions falling into Group C, the most consumptive oriented group. As one might expect, a smaller proportion of the hunter sample falls into Group A, the most nonconsumptive oriented group.

Table 3 shows the range, mean, and median for several variables from the voter survey. The columns indicate information for each of the three attitudinal groups, with absolute numbers of respondents in each group shown in parentheses. The first variable shown is whether respondents have "ever gone on an outing which included wildlife viewing as one of the things you **planned** to do." The proportion responding yes strictly decreases from Group A (82.8%) to Group C (53.2%). The average numbers of primary overnight wildlife viewing trips[5] and wildlife viewing day trips[6] both decrease as one moves from Group A to Group C. The trend is downward from Group A to Group C when one looks at wildlife viewing related equipment purchases ($437, $281, $114, respectively). The equipment purchase numbers are based on responses to the question: "Please tell us about any other equipment or special clothing that you purchased in 1991 for which wildlife viewing was one of the main reasons you purchased the good." Several categories were

---

[5]Overnight wildlife viewing trips were defined by the question: "Did you take any **overnight trips in Alaska** (other than hunting or trapping trips) during 1991 **on which wildlife viewing was one of the activities you planned?** By overnight trip we mean a trip on which you spent one or more nights away from home." Respondents were then led through a process whereby they listed their overnight wildlife viewing trips and classified them as primary or secondary as follows: "People may plan several different activities when they decide to take an overnight trip. Wildlife viewing may be just one of many reasons for a particular trip, or it may be the single most important reason. We will use the terms PRIMARY and SECONDARY to separate trips which were made primarily for wildlife viewing from those where the importance of wildlife viewing was secondary to other activities." followed by: "The best way to determine if a trip is a PRIMARY WILDLIFE VIEWING TRIP is to ask yourself the question 'would I have made this trip to this place if I had not been planning to view wildlife?' If your answer is NO, then that is a PRIMARY WILDLIFE VIEWING TRIP. If your answer is YES, then that is a SECONDARY WILDLIFE VIEWING TRIP." and some examples: (1) "A trip to Denali National Park to specifically find wildlife to watch or photograph is a PRIMARY WILDLIFE VIEWING TRIP; whereas A trip to Denali National Park to specifically photograph Mt. McKinley, but where you watch or photograph any wildlife you happen to see is a SECONDARY WILDLIFE VIEWING TRIP." (2) "A fishing, hiking, boating, sightseeing, or horseback trip to an area you selected because the area offered wildlife viewing is a PRIMARY WILDLIFE VIEWING TRIP. That is, you would have gone to a different area had that area not offered the opportunity to view wildlife; whereas A fishing, hiking, or horseback trip to an area you chose for some reason other than wildlife viewing, but where you also planned to watch wildlife is a SECONDARY WILDLIFE VIEWING TRIP. That is, wildlife viewing was one of your planned activities, but not the one that determined where you went."

[6]Wildlife viewing daytrips were defined by the question: "Did you take any **day trips** in Alaska (other than hunting or trapping trips) in 1991 on which wildlife viewing was **one of the main activities you planned?** By day trip we mean a trip on which you left and returned home in the same day."

provided to help focus respondents' thinking (cameras, lenses, and other photographic equipment; binoculars, scopes, etc.; camping equipment; special clothing; skis or snowshoes; bird feeders or seed; other (please specify)). The proportion of group members who took "an overnight trip in Alaska (other than hunting or trapping trips) **during 1991** on which wildlife viewing was one of the activities you planned" falls moving from Group A to Group C. Those overnight trips include both trips on which wildlife viewing was the primary purpose of the trip and those on which wildlife viewing was a secondary purpose. All the proportions and means in Table 3 discussed to this point are significantly different between groups at the .05 level of significance. The last row of Table 3 shows the proportion of those respondents who took an overnight wildlife viewing trip in 1991 for whom wildlife viewing was the primary purpose of at least one of those trips. That proportion is strictly decreasing as one moves from Group A (57.0%) to Group C (41.4%). The proportions are significantly different between all three groups at the .10 level, and significantly different between Groups A and C at the .05 level.

In both the RV and RH surveys respondents were asked to list the overnight trips they took in Alaska during 1991 and provide basic information about each trip--when the trip occurred, destination, target species. After listing their overnight trips respondents were led through a process in which they randomly selected one trip about which to provide more detailed information. Details of that procedure can be found in McCollum and Miller (1994), but examination of the responses indicates that the vast majority of respondents followed the instructions, and did, in fact, choose the correct trips on which to report. Respondents' randomly selected trips were designated their **"selected trips."** When listing their trips, voters were asked to classify their trips as primary or secondary according to the guidelines discussed in footnote 5. The random selection process in the RV survey was designed so that a primary wildlife viewing trip was chosen as the selected trip if the respondent had any primary trips at all. Only in cases where there were no primary trips did respondents provide detailed information about a secondary trip. Group comparisons pertaining to voters' selected primary wildlife viewing trips are shown in Table 4. The first row shows the comparison (in terms of range, mean, and median) of trip related expenditures. The means range from $567 for Group A to $272 for Group C. The differences in means between all three groups are statistically significant at the .05 level. The net value of the trip was from an open-ended CV question; the CV question from the voter survey is shown in Figure 1. The means--$143 for Group A, $138 for Group B, $88 for Group C--are not significantly different either between all three groups or between the two extreme groups. The last row in Table 4 compares the gross value of the selected trip for all three groups. Gross value was defined as the sum of trip expenditures and net value. The mean gross values range from $726 for Group A to $385 for Group C. The difference between groups A and C is statistically significant at the .05 level and the differences between all three groups are statistically significant at the .10 level.

Information about numbers and types of hunting trips, and hunting related equipment and supply expenditures is presented in Table 5. The mean number of overnight big game or waterfowl hunting trips taken in 1991 ranges from 0.758 for Group A to 1.46 for Group C. Group A members averaged 2.18 big game hunting day trips compared to 4.65 for Group B and 5.10 for Group C; small game hunting day trips averaged 2.60 for Group A, 4.60 for

Group B, and 5.35 for Group C. The numbers of overnight big game hunting trips, big game day trips, and small game day trips are significantly different between all three groups at the .05 level. Waterfowl hunting day trips average about 2 for all three groups, and differences are not statistically significant. Hunting related equipment and supply purchases averaged $441 for Group A, $572 for Group B, and $676 for Group C. Those expenditures are significantly different between all three groups at the .05 level. The proportions of each group taking an overnight big game hunting trip in 1991 (44.7% for Group A, 61.6% for Group B, and 66.0% for Group C) are significantly different between all three groups at the .05 level. The proportions of hunters whose overnight hunting trip was primary does not differ significantly between groups. About 90% of overnight big game and waterfowl hunting trips were primary for all groups.

Table 6 focuses on the selected primary purpose big game or waterfowl hunting trips[7] broken down by attitudinal group. Trip expenditures ranged from $577 for Group A to $805 for Group C, with the differences between those two extreme groups being significant at the .05 level. Net economic values were estimated using an open-ended CV question like the wildlife viewing question shown in Figure 1. The mean values ($137 for Group A, $162 for Group B, $207 for Group C) show the expected progression, but differences are not statistically significant between groups. Gross trip value--again, defined as trip expenditures

[7]The definition of overnight trip for hunting was the same as that for wildlife viewing, i.e., that the person spent one or more nights away from home. Trips were classified as primary based on the response to the question: "Was big game or waterfowl hunting the primary reason for your SELECTED TRIP? By primary we mean you would not have taken the trip had you not been planning to hunt."

plus net value--ranged from \$741 for Group A to \$998 for Group C. The difference in mean between Groups A and C is statistically significant at the .10 level.

Tables 7 and 8 show average gross values placed on day trips to potential wildlife viewing sites for different species by attitudinal group for voters and hunters, respectively. Those values came from highly simplified CV scenarios that asked respondents how often they thought they might take a day trip to visit a site like the ones described if the cost were as specified. The sites were described as: "Suppose an easily accessible wildlife viewing site were available that offered good views of wildlife in natural surroundings from a close, but safe, distance." Then the particular wildlife species were specified along with randomly selected cost amounts. The actual question from one of the surveys is shown in Figure 2. Those responding "at least once" or "more than once" were combined and designated as "yes" responses in a simple logit model with a constant term and the offer amount as independent variables, and probability of a "yes" response as the dependent variable. The estimated distributions, truncated at the .01 level of probability, were then used to calculate mean values. Likelihood ratio tests performed on the estimated distributions--for both RV and RH--showed that the distributions were not the same for all three groups at the .05 level, for any species. Neither were the estimated distributions the same for Groups A and C at the .05 level for any species. Thus, we infer the mean values to be different between groups for each species in Tables 7 and 8.

Interpretation of the estimated distributions for visiting the potential viewing sites is difficult for several reasons, including: the looseness of the scenario specification in terms of site characteristics and access, and the fact that they are open to interpretation by individual respondents; the implied certainty of wildlife viewing--a herd of caribou, several Dall sheep, a large concentration of eagles, etc.; the hypothetical specification of the sites which may or may not be believable to some respondents; and that day trips may or may not be viewed as realistic by some respondents. Nonetheless, the scenarios and responses can be viewed as containing some crude level of information regarding potential demand for new or expanded/enhanced wildlife viewing sites, and on the relative values that Alaskans place on viewing different species.

Browsing Tables 7 and 8, it is interesting that gross values always decrease moving from Group A to Group C, i.e., members of Group C always place less value on the wildlife viewing sites than members of Group B, who always place less value on the wildlife viewing sites than members of Group A. That holds for both the voter and hunter samples. While there is some reordering of species values between groups, it is also interesting that the grizzly bear viewing site always provides the highest level of gross value for both samples, while seabirds and moose are always at the bottom. Comparing Tables 7 and 8, the relative order of values for Group A is virtually the same between the hunter and voter samples. That for Group C is similar between the two samples, except that wolves and marine life/whales are reversed. The most difference between the hunter and voter samples appears in Group B. In all cases, the gross value derived from the voter sample is higher than that derived from the hunter sample. The proportion of Alaskans purchasing a hunting license in 1991 was around 23%, indicating that hunters would be expected to constitute only a minority of the voter sample. That implies hunters, on average, place lower values on the potential wildlife viewing sites than nonhunters.

Discussion and Conclusions

For Alaska voters the observed pattern of behavior, in terms of numbers and types of wildlife viewing trips taken by members of Groups A, B, and C, is consistent with the attitudes expressed by group members on which group classifications were made. Group A, who expressed the most positive attitudes toward wildlife viewing, was the most likely to have ever gone on an outing on which wildlife viewing was a planned activity, and took the most wildlife viewing trips in 1991 in all categories--overnight, primary overnight, and day trips. That same pattern is seen in wildlife viewing related equipment purchases--less spending as one moves from Group A to Group B to Group C. The implication of those observations is that actions by members of the groups bear out their attitudes.

Among those members of the groups who took an overnight trip on which wildlife viewing was the primary activity, net economic value of and expenditures made for the selected trip exhibit a downward trend from Group A to Group C. Again, this pattern is in line with the attitudes on which groups were defined.

For resident hunters, the results again follow the pattern suggested by the attitudes toward hunting and consumptive uses of wildlife. The trend is upwards from Group A to Group C in number of overnight big game hunting trips, and both big game and small game hunting day trips. The amount of annual hunting related equipment purchases increases steadily as one moves from Group A to Group C. All those observations of behavior mirror the attitudes expressed by each group, i.e., that hunters become more consumptive in their attitudes going from Group A to C.

Among those hunters who took overnight big game or waterfowl hunting trips in 1991, Group C had a somewhat higher proportion of trips on which hunting was the primary purpose of the trip. In terms of net value placed on the selected trip, values strictly increase from Group A to Group C. Expenditures while on the selected trip exhibit an upward trend.

For both the hunters and the voters, average values for the potential wildlife viewing sites decreased from Group A to Group C. This is expected based on the Groups' attitudes toward wildlife viewing and consumptive uses of wildlife.

What these results show is that attitudes expressed by the survey respondents which were used to classify them into groups carry through to the behavioral intentions expressed in the responses to the contingent valuations of the selected trip and the potential wildlife viewing sites. Those attitudes also carry through to actual behavior expressed in the number and types of wildlife related trips--both hunting and viewing--and in the hunting or wildlife viewing related equipment purchases, and in the expenditures on the selected trip. Further, the discriminant function estimated using data from the voter survey appears to have done a good job at categorizing respondents to the resident hunter survey. The behavioral intentions and actual behavior expressed by the resident hunters are consistent with the attitudes by which they were categorized. Results from this study would indicate that a behavioral intention, namely contingent valuation, can reliably portray preferences that can be reflected in actual behavior.

References

Fishbein, M., I. Ajzen. 1975. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley Publishing Co.

McCollum, D.W., S.M. Miller. 1994. Alaska Voters, Alaska Hunters, and Alaska Nonresident Hunters: Their Wildlife Related Trip Characteristics and Economics. Anchorage: Alaska Department of Fish and Game, Division of Wildlife Conservation.

Miller, S.M., R. Sinnott, D.W. McCollum. 1994. Differentiating among Wildlife-related Attitudinal Groups in Alaska. Transactions of the 59th North American Wildlife and Natural Resources Conference, 59:512-523.

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

Table 1. Attitudinal Statements Used to Define Groups

---

* **I am interested in knowing more about how to find and watch wildlife.**

* **I would probably stop or slow down to look for wildlife if I saw a sign along the highway indicating good wildlife viewing.**

* **I think more concern should be given to protecting the land and water where wildlife live.**

In general, I approve of hunting wildlife for meat.

* **In general, I approve of hunting wildlife for trophies.**

* **In general, I approve of trapping wildlife.**

* **I think hunters have too much influence on wildlife management.**

* **I think environmentalists have too much influence on wildlife management.**

* **I think people living outside Alaska have too much influence on wildlife management in Alaska.**

* **I think more areas in the state should be managed and developed for wildlife viewing.**

* **I think more areas in the state should be managed and developed for wildlife viewing, even if that means closing some areas to hunting.**

* **In general, I believe it is more difficult to see wild animals in areas where those same animals are hunted than in areas where they are not hunted.**

I believe more areas in the state should be closed to hunting.

I like to eat game meat.

* **I like to go sport fishing.**

* **I support killing wolves in some areas of Alaska to increase the numbers of moose and caribou.**

---

* and bold face type indicate the 13 questions that were asked on both the resident voter and resident hunter surveys, and were used to estimate a discriminant function, based on the voter data, that was then used to classify resident hunters into the same groups.

Table 2. Relative Proportions of Samples Falling into Each Attitudinal Group

	Voter Sample	Resident Hunter Sample
Group A	20.7%	9.2%
Group B	59.9%	73.4%
Group C	19.4%	17.3%

Table 3. Resident Voters Wildlife Viewing Related Behaviors by Group

	Group A (n = 444)	Group B (n = 1,283)	Group C (n = 415)
Wildlife Viewing Trip Ever			
Yes[a]	82.8%	73.1%	53.2%
Number of Primary Wildlife Viewing Trips			
Range	0 - 6	0 - 7	0 - 6
Mean[a]	0.430	0.305	0.135
Median	0	0	0
Wildlife Viewing Day Trips			
Range	0 - 60	0 - 90	0 - 80
Mean[a]	4.60	4.19	2.16
Median	2.0	1.0	0
Wildlife Viewing Related Equipment Purchases			
Range	0 - $45,230	0 - $20,405	0 - $3,775
Mean[a]	$437	$281	$114
Median	0	0	0
Overnight Wildlife Viewing Trips in 1991			
Yes[a]	51.1%	45.6%	32.1%
Percent Taking Overnight Trips in 1991--Primary Purpose for Wildlife Viewing[b]	57.0%	49.2%	41.4%

[a]Means or proportions are significantly different between all three groups at the .05 level.

[b]Proportions are significantly different between Groups A and C at the .05 level, and between all three groups at the .10 level.

Table 4. Resident Voters Wildlife Viewing Related Behaviors by Group--
Primary Wildlife Viewing Trips

	Group A (n = 90)	Group B (n = 176)	Group C (n = 24)
Selected Trip-- Trip Purchases			
Range	$0 - $3,116	$0 - $3,750	$0 - $1,420
Mean[a]	$567	$476	$272
Median	$292	$287	$182
Selected Trip-- Net Value			
Range	$0 - $2,000	$0 - $1,000	$0 - $500
Mean	$143	$138	$88
Median	$100	$100	$50
Selected Trip-- Gross Value			
Range	$0 - $5,116	$0 - $4,750	$50 - $1,420
Mean[b]	$726	$614	$385
Median	$425	$412	$300

[a]Means are significantly different between all three groups at the .05 level.

[b]Means are significantly different between Groups A and C at the .05 level, and between all three groups at the .10 level.

Table 5. Resident Hunters Hunting Related Behaviors by Group

	Group A (n = 192)	Group B (n = 1,526)	Group C (n = 350)
Number of Overnight Big Game Hunting Trips			
Range	0 - 10	0 - 10	0 - 10
Mean[a]	0.758	1.25	1.46
Median	0	1.0	1.0
Big Game Hunting Day Trips			
Range	0 - 21	0 - 60	0 - 60
Mean[a]	2.18	4.65	5.10
Median	0	2.0	2.0
Small Game Hunting Day Trips			
Range	0 - 30	0 - 70	0 - 60
Mean[a]	2.60	4.60	5.35
Median	0	1.0	2.0
Waterfowl Hunting Day Trips			
Range	0 - 45	0 - 60	0 - 40
Mean	1.77	1.99	1.80
Median	0	0	0
Hunting Related Equipment Purchases			
Range	$0 - $7,500	$0 - $12,000	$0 - $15,040
Mean[a]	$441	$572	$676
Median	$0	$156	$130
Overnight Big Game Hunting Trips in 1991			
Yes[a]	44.7%	61.6%	66.0%
Percent Overnight Trips in 1991--Primary Purpose for Big Game Hunting	89.3%	90.7%	92.6%

[a]Means or proportions are significantly different between all three groups at the .05 level.

Table 6. Resident Hunters Hunting Related Behaviors by Group--
Primary Big Game Hunting Trips

	Group A (n = 75)	Group B (n = 789)	Group C (n = 212)
Selected Trip-- Trip Purchases			
Range	$0 - $2,969	$0 - $17,612	$0 - $7,703
Mean[a]	$577	$715	$805
Median	$340	$412	$430
Selected Trip-- Net Value			
Range	$0 - $2,000	$0 - $2,000	$0 - $5,000
Mean	$137	$162	$207
Median	$100	$100	$75
Selected Trip-- Gross Value			
Range	$58 - $3,237	$0 - $6,920	$0 - $8,103
Mean[b]	$741	$858	$998
Median	$493	$553	$633

[a]Means are significantly different between Groups A and C at the .05 level.

[b]Means are significantly different between Groups A and C at the .10 level.

Table 7. Average Gross Value Placed on Potential Wildlife Viewing Sites for Different Species by Resident Voters by Group

	Group A (n = 444)	Group B (n = 1,283)	Group C (n = 415)
Caribou[a]	$357	$306	$125
Moose[a]	$171	$129	$ 66
Wolves[a]	$423	$295	$165
Sheep[a]	$271	$240	$101
Whales[a]	$366	$297	$130
Bears[a]	$565	$453	$221
Seabirds[a]	$222	$145	$ 43
Eagles[a]	$368	$260	$ 85

[a]Likelihood ratio tests showed: (1) estimated distributions are not the same for all three groups at the .05 level, (2) estimated distributions are not the same for Groups A and C at the .05 level. Hence, we infer the mean values to be different between groups.

Table 8. Average Gross Value Placed on Potential Wildlife Viewing Sites for Different Species by Resident Hunters by Group

	Group A (n = 192)	Group B (n = 1,526)	Group C (n = 359)
Caribou[a]	$271	$176	$ 68
Moose[a]	$101	$ 78	$ 43
Wolves[a]	$391	$240	$ 86
Sheep[a]	$199	$162	$ 66
Whales[a]	$271	$206	$ 94
Bears[a]	$525	$358	$211
Seabirds[a]	$169	$ 86	$ 31
Eagles[a]	$272	$154	$ 59

[a]Likelihood ratio tests showed: (1) estimated distributions are not the same for all three groups at the .05 level, (2) estimated distributions are not the same for Groups A and C at the .05 level. Hence, we infer the mean values to be different between groups.

Figure 1. The Contingent Valuation Question

The next set of questions will ask you how satisfied you were with your SELECTED TRIP and the money you paid for it.

25. All things considered--the expenses, the wildlife, whether you saw the particular kinds of wildlife you were looking for, the overall quality of the trip--do you feel that your SELECTED TRIP was worth the money you paid? If you had it to do all over again, would you take that exact same wildlife viewing trip for the same cost? *(CIRCLE YOUR ANSWER)*

NO *Please go to the next page, PART THREE*

YES *Please go on to Question 26*

26. The expenses for your SELECTED TRIP could have been higher or lower. For example, transportation and other expenses rise and fall over time. Suppose your costs had been higher. How much could the cost of your SELECTED TRIP have increased before you would have decided it was just not worth it and you would not have taken your SELECTED TRIP?

In other words: Up to how much more, in addition to what you actually paid for your SELECTED TRIP, would you have paid to take your SELECTED TRIP?

I would have paid as much as $__________ more than I actually paid to take my SELECTED TRIP.

27. If you answered $0 to the previous question, please tell us what that means. *(CIRCLE ONE NUMBER)*

1 **I would not have paid any more to take my SELECTED TRIP. That trip was worth exactly what I paid and no more.**

2 **I answered $0 because I could not put a number on how much more I would have paid for my SELECTED TRIP.**

3 **I answered $0 for other reasons. Please specify______________________**

Figure 2. Species Viewing Valuation Scenarios

How interested are you in new sites to view wildlife? Suppose an easily accessible wildlife viewing site were available that offered good views of wildlife in natural surroundings from a close, but safe, distance. Please tell us how often you think you would make a day trip to visit sites like the following over the next five years.

Each question indicates a cost per person to make the round trip. That cost is different in every questionnaire. Please give us your answer even if the amount seems ridiculously high or low so that we can consider a full range of values. *(CIRCLE ONE NUMBER FOR EACH STATEMENT)*

	Never	At Least Once	More Than Once
How often do you think you would take a day trip to visit a site over the next five years where you could expect to see....			
A herd of caribou if it cost $255 to make the trip?	0	1	2
Moose if it cost $290 to make the trip?	0	1	2
A pack of wolves either from the ground or from an airplane if it cost $55 to make the trip?	0	1	2
Several Dall sheep if it cost $370 to make the trip?	0	1	2
Typical marine life, and about half the time you would see whales, if it cost $25 to make the trip?	0	1	2
A large concentration of grizzly bears if it cost $100 to make the trip?	0	1	2
A large concentration of sea birds if it cost $500 to make the trip?	0	1	2
A large concentration of eagles if it cost $140 to make the trip?	0	1	2

Note: The ranges of offers were:
$1 to $1,000 for wolves and grizzly bears.
$1 to $900 for caribou, Dall sheep, and eagles.
$1 to $600 for moose, marine life/whales, and sea birds.

STATED AND REVEALED PREFERENCES, BENEFIT TRANSFER, AND WELFARE DISTRIBUTION EFFECTS OF VALLEY RANCH OPEN SPACE

Richard G. Walsh, John R. McKean, and Randall S. Rosenberger
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, CO 80523-1172

ABSTRACT

A national study of household preferences for preservation of a mixed private-public good, valley ranch open space, tested whether information on revealed preference would influence stated preference. Asking questions about direct and indirect use prior to willingness to pay questions resulted in approximately a 50 percent reduction in estimated willingness to pay, consistent with the NOAA recommendation. However, the results suggest that (1) transfer of estimated preservation benefits based solely on stated preference would be understated in this case, and (2) resource allocation based on stated preference would be inequitable since it favors higher income households while revealed preference for the resource by lower income households is greater than higher income households. A more equitable distribution of the benefits of preserving valley ranch open space would result if the allocation were based on some combination of <u>stated and revealed</u> preferences, accounting for both higher income household ability to pay and lower income household ability to allocate own time resources to direct and indirect use. These results support the work by Hanemann (1991) with respect to the divergence between willingness to pay and willing to accept compensation related to the welfare effect of in-kind services of public goods.

INTRODUCTION[1]

We designed a national household survey to test whether reminders, in the form of questions on direct and indirect uses prior to the willingness to pay (WTP) questions, would influence contingent valuation method (CVM) estimates of WTP total value. In other words, would information on revealed preference influence stated preference (Randall et al. 1990). The context of this experiment is a CVM survey of the national benefits from preservation of ranch open space wherever it remains in river valleys throughout the country. We found that two versions of the mail survey, differing only by inclusion of the reminder, did yield statistically different modified payment card WTP estimates for two randomly drawn subsamples from the same sample frame. The reminder coefficient in a pooled aggregate benefit function reduced WTP by 45 percent, all else equal. Also, the group mean WTP fell by 56 percent with the reminder, from $51 to $23 per household, while average income was not significantly different for the two versions, $42,200 vs. $40,700.

This may provide some support for two recommendations of the NOAA panel on CVM, that respondents be explicitly reminded about substitutes and budget constraints prior to answering WTP questions, and that WTP estimates be reduced by 50 percent (Arrow, et al. 1993:4608-4609). But NOAA observers should not read too much into the results. A recent test of reminding Oregon respondents about substitutes and budget constraints found no significant difference in a dichotomous choice logit equation for benefits from reducing fire hazards by half on 3 million acres of old growth forest in the state (Loomis, et al. 1994). This

---

[1] This study was funded, in part, by the Routt County Board of Commissioners, and the Colorado Agricultural Experiment Station, Project W-133, Benefits and Costs Transfer in Natural Resource Planning, Colorado State University, Fort Collins. We are grateful for the assistance of the Garden Park Paleontology Society, and several individuals especially Dan Grenard, Donna Engard, Julie Schaefers, Kun John, Wayde Morse, Ken Bonetti, Bob Young, and John Loomis.

emphasizes the importance of replicating empirical tests of recommendations for improvement in CVM procedures.

The difference between these two tests may be in the cognitive effect of revealed preference questions compared to short descriptions of possible substitutes and budget constraints. To ask respondents about possible direct and indirect use introduces detailed information on the money and time price of closely related goods or activities which are usually complements rather than substitutes, and it adds a separate time constraint to the traditional budget constraint (Bockstael and McConnell, 1981; Bockstael, et al. 1987). The economic analysis of recreation demand for direct and indirect use of valley ranch resources is based on the theory of nonmarket production and consumption activities (Becker, 1965; 1976). Individuals and households produce and consume a recreation experience by combining inputs of their own time, purchased goods and services, environmental resources, and other inputs (skill, knowledge, etc.). They will attempt to allocate direct and indirect use of valley ranch resources to get the most enjoyment they can from the experience subject to constraints of income, time, and other conditions (Randall and Stoll, 1983). This means that valley ranch resources are intermediate goods or inputs whose demand is derived from the final demand for the experience of direct and indirect recreation use.

The objective of this paper is to explore the possibility of benefit transfer of a national WTP study to estimate the value of county and state ranch open space programs. The hypothesis is that benefit transfer of WTP values equal benefit transfer of direct and indirect use values. Our approach is to estimate statistical functions for the three different measures of value and to test whether the coefficients from the WTP benefit function are similar enough to the coefficients from recreation demand functions for direct and indirect use to support benefit transfer of WTP studies.

In current recommendations for improvement in CVM procedures, it is assumed that we can measure the total value of direct and indirect uses and services of the resource with a single dichotomous choice WTP question (Arrow, et al. 1993). This approach has become standard CVM practice because it is consistent with economic theory and has other obvious advantages. A single WTP question is more cost-effective than asking detailed questions about direct and indirect uses, and avoids possible double counting when summing consumer surplus of direct and indirect uses to obtain total value. Attempts to identify direct and indirect uses by allocation of total WTP have been challenged as not consistent with economic theory that defines onsite use, option, existence, and bequest demands as tastes rather than useful economic services.

A possible second CVM question on willingness to accept (WTA) compensation for the loss of economic services of environmental resources is not recommended as a valid procedure (Arrow, et al. 1993). In empirical tests, many households refuse to sell or give exaggerated estimates of WTA compensation. As a result, expressed WTA is usually several orders of magnitude greater than WTP, which is not consistent with the economic theory of equality between the two values, with negligible income effects. However, recent work by Hanemann (1991) suggests that the difference in WTP and WTA compensation for a change in the quantity of a public good depends not only on an income effect but also a substitution effect. The smaller the substitution effect (i.e. the fewer privately marketed substitutes available for the public good) the greater the disparity between WTP and WTA. Higher expressed WTA compensation may be due to the fact that direct and indirect use significantly increases the total or full income of households, including market plus <u>in-kind services</u>. Where CVM questions on WTA are not recommended as a valid procedure, consumer surplus of current direct and indirect use may be an effective economic indicator of WTA compensation. The measure could be inclusive if current users and nonusers were asked to report frequency of past and expected future direct and indirect use. Valley ranch open space may be a form of mixed private-public natural capital that is non-substitutable. If this is so, it would seem to be important to get a better understanding of the sensitivity of consumer surplus of in-kind services to the level of market income.

The national perspective provides an opportunity to develop demand functions and benefits for two curiously neglected uses, even though they are, by far, the largest nonmarket uses of recreational resources (Walsh, 1986; Walsh, et al. 1990). They are the outdoor recreation activities of sightseeing or pleasure driving, and the indoor recreation activities of reading, watching programs, etc. about resource-related subjects, as suggested by Larson (1993). The resource has several possible beneficial effects that are not fully reflected in its market value. Public nonmarket uses and services include: floodplain protection that reduces property damage and loss of life; water quality and wetland protection for fish and wildlife habitat; open space greenways along scenic highways, byways, riverways, and trails linking communities; and the resource base for the continued development of western ranch culture, music, fiction, art, clothing, furniture, etc.

County commissioners and others responsible for land use programs need to know what the public benefit would be from the protection of open space in specific valley ranch areas by purchasing development rights and other means. The CVM results of five regional case studies (Table 1) suggest that most local residents would favor preservation of farm and ranch open space and would be willing to pay for it. Because of the low population in rural counties, the

Table 1. Studies of Household Willingness to Pay for Preservation of Farm and Ranch Open Space in Canada and the United States, 1994 Dollars.

Study Area, Resource, and Source	Total Quantity, Acres and Percent	Average Annual Willingness to Pay per Household	Investment Value per Acre at 6% Interest[b]	Year, Sample, and Population	Type of Survey and Valuation Format
SOUTH					
Prime agricultural land in Piedmont area of Greenville County, South Carolina. Bergstrom, Dillman and Stoll (1985)	18,000 (25%) 36,000 (50%) 54,000 (75%) 72,000 (100%)	$9 $10 $12 $14	$902 $526 $401 $351	1982, 250 households in Greenville county, with 108,193 households	Mail survey, CVM open ended question. WTP: (1) increase in county taxes; or (2) donation to private county prime land conservation fund, photos.
ALASKA					
Old Colony and Homestead farms in the Matanuska-Susitna valleys near Anchorage in southcentral Alaska. Beasley, Workman and Williams (1986)	3,500 (50%)[a] 7,000 (100%)	$114 $216	$4,831 $4,577	1983, 119 households in the Palmer, Wassilla and outlying areas, with 8,900 households	Personal interviews, CVM, interative bidding, with WTP choice of: (1) sales tax, (2) property tax, or (3) donation to private, farmland preservation fund, photos.
NORTH					
Farms in Deerfield, East Longmeadow, and Greenfield townships in western Massachusetts. Foster, Halstead and Stevens (1982); Halstead (1994)	1,322 (33%)[a] 2,644 (66%)[a] 3,967 (100%)	$200 $291 $358	$12,279 $8,933 $7,325	1981, 85 households in three townships, with 4,870 households per township	Personal interviews, CVM, iterative bidding, choice of: (1) local sales tax; (2) state income tax; or (3) donation to local private farm land preservation fund, photos.
CANADA					
Agricultural land in the Kent, Albert, and Westmorland three-county area of New Brunswick province in eastern Canada. Bowker and Didychuk (1994)	23,750 (25%) 47,000 (50%) 71,250 (75%) 95,000 (100%)	$49 $68 $78 $86	$1,195 $838 $634 $524	1991, 93 households in the three counties, with 34,740 households	Personal interviews, CVM, modified payment card, donation to private non-profit conservation foundation.

Table 1. Studies of Household Willingness to Pay for Preservation of Farm and Ranch Open Space in Canada and the United States, 1994 Dollars (Continued...)

Study Area, Resource, and Source	Total Quantity, Acres and Percent	Average Annual Willingness to Pay per Household	Investment Value per Acre at 6% Interest[b]	Year, Sample, and Population	Type of Survey and Valuation Format
ROCKY MOUNTAINS	Steamboat				
Ranch open space in the	2,500 (25%)	$42	$1,400	1994, 171 registered	Mail and phone, CVM
Yampa River valley	5,000 (50%)	$70	$1,167	voters in Routt County,	modified payment card
around Steamboat Springs	7,500 (75%)	$80	$889	with 6,200 households	and behavior questions,
resort and elsewhere in	10,000 (100%)	$74	$617		WTP in taxes or prices
Routt County, Colorado.					of goods and services.
Rosenberger, Walsh,	Elsewhere				
McKean, and Mucklow	10,000 (25%)	$54	$450		
(1995)	20,000 (50%)	$98	$408		
	30,000 (75%)	$118	$328		
	40,000 (100%)	$75	$156		
ROCKY MOUNTAINS					
Ranch open space in the	10,000 (100%)	TCM		1993, 403 households, on	Personal interviews,
Yampa River valley		$12	$7,900	summer vacation in	CVM open ended and
around Steamboat Springs		CVM		Steamboat Springs, with	TCM behavior questions.
resort in Routt County,		$16	$10,500	394,000 summer vacation	WTP travel costs.
Colorado. Walsh,		(per user day		user days	
McKean, Rosenberger and		sightseeing)[c]			
Mucklow (1994)					
UNITED STATES					
Ranches and livestock	12 m (25%)[d]	$61	$8,133	1994, 512 households in	Mail and phone, CVM
farms in river valleys of	24 m (50%)	$95	$6,333	U.S., with 96 million	modified payment card,
50 states, total 48 million	36 m (75%)	$108	$4,800	households	open ended, and behavior
acres or 2% of land usage	48 m (100%)	$102	$3,400		questions. WTP taxes
in U.S. Walsh, McKean,					and volunteer time.
and Rosenberger (1995)	Direct use	$386	$12,880		
	Indirect use	$213	$7,120		
	50% increase in direct and indirect use	$414	$13,780		

[a] Approximated. [b] Investment value is assumed equal to present value: average annual willingness to pay per household multiplied by total households, discounted at 6 percent interest in perpetuity, and divided by acres. [c] A negative binomial demand function estimated that seeing valley ranch open space accounted for an average of approximately 22 percent of $68 per user day of consumer surplus from recreation sightseeing trips. [d] A quadratic benefit function adjusted for the proportion of households expressing a preference for at least these amounts, includes WTP per household and reported willingness to volunteer time valued at the reported wage rate (average $14) for preservation of valley ranch areas.

opportunity cost of development rights often exceeds local willingness to pay. As a result, although many counties and states have established programs to protect farm and ranch open space, typically very little county or state tax money is spent on preservation programs (Lopez, et al. 1994). However, the value of open space in rural counties is likely to be conservative because it does not include possible benefits to tourists who visit the county and the general public who do not, both of which result in the "aggregation" problem discussed in guidelines to CVM nonmarket valuation research (Mitchell and Carson, 1989). Two studies summarize our attempts to begin evaluating the benefits of these two groups.

It is estimated that valley ranch livestock growers occupy 48 million acres or about 2 percent of the land in the United States. According to a recent study, nearly 300,000 acres change to other uses each year (U.S. Dept. of Agriculture, 1985). If present trends continue, over one-third of the remaining valley ranch area will be lost by the year 2040. Increasing scarcity of the resource in some rural communities has led county commissioners to propose studies of the market and nonmarket value of remaining valley ranch land to assess whether its preservation would contribute to public welfare. In a balanced approach to land use management, some valley ranch land would be open to carefully planned development while some would be best suited for preservation as open space. Nations around the world face similar problems of estimating how much they can afford to pay for the protection of farm and ranch open space in river valleys (Willis and Whitby, 1985; Young and Allen, 1986). The possibility of expanded development for other uses and the accompanying probability of damage to the quality of open space provides a realistic setting for investigating the empirical significance of its preservation value. The purpose is to develop and apply cost-effective research procedures to estimate statistical demand and benefit functions for valley ranch resources which more nearly approach the goal of including total value to society. The study contributes to the development of practicable methodology for application of economics to valuation of the resource. The survey provides new information on several economic indicators of benefits. The results include some of the information necessary to compare the benefits of increments in valley ranch protection with costs.

RESEARCH PROCEDURE

Data for this study are from a mail survey of 512 households completed by a national market research firm in the first quarter of 1994. The useable response rate was 51 percent of the national sample of 1,000 households. The sample frame was a consumer panel stratified to represent U.S. Census household characteristics. Use of the approach in this study is based on a recommendation by the federal advisory panel on contingent valuation research (Arrow, et al. 1993) that successful commercial market survey methods be introduced in nonmarket resource economic studies. The basic panel sample stratification by region may be a more cost-effective way (at about $12 per case or $6,000 for 512 cases) to represent the U.S. population than personal interviews with random cluster samples from a few representative communities (Carson, et al. 1992). The objective of both methods is to approach as nearly as possible the characteristics of a true random probability sample, which is a statistical ideal beyond the reach of social science research.

Any bias introduced by over or under sampling can be reduced by substituting the correct sample proportion in the statistical regression, unless the households sampled have more interest in the subject than the population (Mitchell and Carson, 1989). The study included two tests for possible bias related to level of interest in the subject, as recommended by guidelines to recreation and environmental economic research (U.S. Water Resources Council, 1983). Ten percent of the nonrespondents were interviewed by phone and a sample of 137 U.S. households were interviewed by random digit dialing. Interest in the phone surveys was significantly higher than for the mail survey, which suggests it provides a conservative estimate of general public interest in the subject.

The sample households were stratified to match the latest U.S. Census percentage distribution within each geographic region according to city size, age of household head, annual household income, and size of household. The sample households are closely representative of U.S. households with respect to the proportion of family and nonfamily households, marital status, type of dwelling, home ownership, education, employment, and occupation. The sample frame, selected for consumer market research, over-sampled women who may account for more household purchase decisions. Whether the national sample accurately represents U.S. Census households is not as important as comparison of the stated and revealed preference subsamples, which are not significantly different with respect to wage rate, education, age, and other demographic variables.

The questionnaire with 27 multi-part questions was designed for clarity and ease of answering (Dillman, 1978). Alternative questions were pretested on three samples of 25-100 households. Figure 1 reproduces the format with reminders of direct and indirect use before WTP questions. This was followed by an open-ended question asking respondents to explain the reasons for answers to the valuation questions. Protest responses of 2.4 percent were included in the analysis. The questionnaire was printed on good quality paper, photo-reproduced, visually uncluttered, and bound in booklet form on legal size paper. Artistic reproductions of the resources were shown at several places throughout the questionnaire and

Please read carefully the definitions listed below and answer the following questions for your household.

Ranch Areas Defined

Scenic valley open space, hay meadow and pasture with wildflowers, birds, deer, other wildlife, grazing cattle, sheep, and horses, corrals and ranch buildings, working ranch hands, cowboys, the basis of western culture, music, fiction, art, clothing, furniture, etc. People enjoy scenic vistas, observing livestock, wildlife, other sight-seeing, bicycling, horseback riding, walking, fishing, hunting, etc. Valley ranches occupy about two percent of the land. In many valleys, this important part of our national heritage is changing to urban and resort uses.

9. Please identify approximately how often during a typical month you participated in the following activities related to RANCHES. About how many hours is this, and how much does it cost?

Your RANCH Related Activities:	How Often In Typical Month (# Of Times)	Total Hours In Typical Month	Expenses In Typical Month (Dollars)
Your time in conversations, seeing pictures, reading, watching television, thinking, etc. about ranch subjects	_____	_____ hours	*(Magazines, books, movie ticket, video rental, etc.)* $_________
Your time on trips actually seeing ranch hay meadows with wildflowers, birds, cattle, sheep, horses, wildlife, etc.	_____	_____ hours	*(Fees, transportation, food, lodging, film, etc.)* $_________

10. Suppose your opportunities to see new and unique ranches and livestock increased by one-half. Would these changes cause you to spend more time, less time, or about the same time in Ranch related activities mentioned in Question 9? **(Check ONE Box AND Record Hours If More Or Less)**

1 ☐ More → How much more? _____ more hours per month
2 ☐ About the same time
3 ☐ Less → How much less? _____ less hours per month

14. How much of these sites listed below do you believe should be preserved? **Check ONE Box**

	None	25% Of Existing Amount	50% Of Existing Amount	75% Of Existing Amount	100% Of Existing Amount	More Than 100% (Increased Through Restoration) (Write In)	Not Sure
Ranches (currently 2% of U.S. land usage)	1☐	2☐	3☐	4☐	5☐	_____%	6☐

15. Would you be willing to pay a proportionate share of the costs of preserving these sites at the level you reported in Question 14 above? In other words, if an election were held today, would you vote "Yes" or "No" on preserving these sites if it cost you $1.00 per year in taxes? **Check ONE Box**

	Yes	No	Not Sure
Ranches	1☐	2☐	3☐

16. This question is hypothetical and intended to provide an economic measure of how much these sites (reported in Question 14) are worth to you. Please estimate the maximum annual amount of money you would pay to preserve them. Assume this is the only way to prevent their changing to other uses. **Check ONE Amount**

	$0	$1	$5	$10	$20	$30	$40	$50	$60	$70	$80	$90	$100	$200	$300	$400	$500	$750	$1,000	Other (Specify)	Not Sure
Ranches ...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	_____	☐

17. Some people prefer or are more able to volunteer their time. In addition to the dollar value you reported (in Question 16 above), what is the maximum number of hours per year you would volunteer to work for the preservation of these sites? **Write In Number Of Hours** **If No Hours, Write In "0"**

	HOURS Per Year	Not Sure
Ranches	_______ hours	6☐

Figure 1. Mail Survey Format With Reminder of Direct and Indirect Use Before Willingness to Pay Questions, Condensed

on the cover letter. The letter was addressed to each individual by name, signed by the survey firm's project leader, and was designed to motivate respondents by explaining the usefulness of the research to recreation planning and the importance of participating in the study. The sponsoring agencies were not identified to avoid possibly influencing respondents. Respondents were asked to consider the value of preserving valley ranch open space, natural areas, and fossil sites, in the same questionnaire. Including questions on public natural areas and fossil discovery sites is consistent with the model developed by McConnell (1989) for estimating the optimal quantity of land in open space, both private and public.

STATISTICAL RESULTS

Three regression models were estimated for WTP and demand for direct and indirect use. Smith and Kaoru (1990) show that for comparability of statistical equations, it is necessary to hold choice of model constant across data sets. Since the truncated Poisson model is recommended for recreation trip demand functions, it also is used for the other two functions. While it is believed to be equally suitable for analysis of demand for occasions of indirect use, it is not necessary to apply the Poisson model in WTP analysis. To do so may affect the significance of variables in the WTP equation where the amount of variation explained (.13) is lower than the demand equations for direct use (.37) and indirect use (.62). There is some heteroskedasticity in the equations for WTP and demand for indirect use which usually does not significantly affect the welfare estimate with a large dependent variable (Creel and Loomis, 1990). Thus, the reminder coefficient in the WTP equation reduced WTP by 45 percent which is somewhat conservative compared to the group mean WTP which fell by 56 percent with the reminder. Still, some bias in the estimate of the coefficients for explanatory variables may be present in two of the three equations, which represents a necessary trade-off to avoid the possibility of introducing a much greater bias in comparison of the three equations without holding the choice of model constant.

The findings shown in Table 2 are consistent with the theory of household production and consumption of nonmarket activities (Bockstael and McConnell, 1981; Bockstael, et al. 1987). Travel cost per trip has a negative coefficient, indicating a downward sloping demand curve for trips. As necessary time per trip increases, opportunity costs rise and households take fewer trips (McConnell, 1992; McKean, et al. 1995). The model represents a variation of the travel cost demand method suggested by Cordell and Bergstrom (1991) where the dependent variable, trips, is specified as annual visits to all sites around a central location, i.e. representative household in a community. The Poisson regression model is appropriate for dependent variables with count (integer) data, such as trips, where each case is a member of a set of positive whole numbers (no fractions or partial trips). A regression-based test (Cameron and Trivedi, 1990) shows no significant overdispersion in the Poisson model.

Consumer surplus was estimated using the equations shown in standard guidelines (Adamowicz, et al. 1989). Using equation 2a (p. 416) for the Poisson semilog functional form, consumer surplus per trip is given by minus one divided by the regression coefficient for cost per trip or price. The direct use model was re-estimated with fewer social economic variables, resulting in a surplus of -1/-0.011 = \$90.91 per household trip or about \$35 per person. The estimate is considered reasonable based on previous research. Six estimates of the consumer surplus of sightseeing outdoor recreation trips in the U.S. average \$26 per recreation day and range from \$13 to \$41, adjusted to 1994 dollars (Walsh, et al. 1990). Average individual consumer surplus for sightseeing and pleasure driving in the U.S. is estimated as \$34 to \$69 per

Table 2. **Household Demand Function for Recreation Trips to See Valley Ranch Areas, United States, 1994.**[a]

Variable	Mean	Coefficient	Elasticity
Travel Cost per Trip, Dollars	33.64	-0.0033	-0.1110
Time per Trip, Hours	4.70	-0.1782	-0.8375
Site Quality Index, Miles	384.08	-0.0005	-0.1920
Leisure Time Resource, Days per Year	99.80	0.0002	0.0200
Opportunity Cost of Time, per Hour	13.32	-0.0192	-0.2557
Household Income	37.96	-0.0081	-0.3075
Valley Ranch Area Protection, Percent	86.15	0.0039	0.3360
Ever Seen a Ranch Area, 0-1	0.71	1.0829	0.7688
Interest in Visiting a Ranch Area, 1-3	2.50	-0.0354	-0.0885
Resident of Western Region of U.S., 0-1	0.23	-0.2966	-0.0682
City Size, over 2 million, 0-1	0.35	-0.7204	-0.2521
Education, Years	14.24	0.1229	1.7501
Household size, Persons	2.51	0.0246	0.0617
Age, years	45.58	0.0047	0.2142
Female respondent, 0-1	0.83	0.5571	0.4624
Black race, 0-1	0.01	7.9109	0.0791
Spanish heritage, 0-1	0.02	1.8759	0.0375
Combined question format, 0-1	0.33	-0.4431	-0.1462
Constant		1.0178	

[a] Adjusted R^2, 0.37; Cases, 92 participants; Dependent variable, logged number of trips per year, mean, 65.3. All variables are significant at the 0.01 level or better except household size, 0.17, leisure time resource, 0.18, and interest in visiting a ranch, 0.24.

trip, adjusted to 1994 dollars (Bergstrom and Cordell, 1991). Consumer surplus of $91 per trip multiplied by an average of 19.3 trips per year equals $1,756 per household, which is adjusted by 0.22 to $386 for the proportion of an average sightseeing trip attributed to seeing ranch open space (Walsh, et al. 1990; 1994).

Table 3 shows a household demand function for indirect recreation use of valley ranch resources, where the dependent variable to be explained is the annual number of occasions of indirect use reported by households who participate in indirect use. The negative coefficient for price or expense per occasion is consistent with the theory of diminishing marginal utility reflected in a downward sloping demand curve for occasions of indirect use. A time per occasion variable is included to account for the exogenous effect of fixed participation (article length, program or movie running time) on demand for indirect use. The frequency of indirect use is positively related to first hand knowledge of valley ranch areas from having seen them in the past and current annual hours devoted to valley ranch-related trips, consistent with a complementary relationship. Although the effect of substitutes for direct use on trips is well known (Rosenthal, 1987), little attention has been given to the effect of complements. For the indirect use model, surplus is -1/-0.0789 = $12.67 per household occasion or about $5 per person per occasion. Consumer surplus of $13 per occasion multiplied by 27.3 occasions per year equals $355 per household, which is adjusted by 0.60 to $213 for household perception of the proportion of the experience attributed to valley ranch resources rather than the artistic contribution of the producer (Walsh and McKean, 1992; 13).

Table 4 shows the household benefit function, where the dependent variable to be explained is WTP per year for households reporting a positive value. WTP increases with level of knowledge from current direct use on trips to see valley ranches and indirect use such as

Table 3. Household Demand Function for Participants in Indirect Recreation Use of Valley Ranch Resources, United States, 1994[a]

Variable	Mean	Coefficient	Elasticity
Cost per Occasion, Dollars	1.69	-0.0789	-0.1333
Time per Occasion, Hours	1.76	0.0255	0.0449
Expenses for Ranch-related Merchandise, Dollars per Year	16.48	0.0052	0.0857
Leisure Time Resource, Days per Year	77.08	0.0010	0.0771
Opportunity Cost of Time, Hourly Wage	13.06	-0.0176	-0.2298
Household Income, Thousand Dollars	39.14	0.0001	0.0031
Site Quality Index, Miles	388.99	-0.0003	-0.1167
Interest in Future Trips, 1-3	2.19	-0.1118	-0.2448
Ever Seen a Ranch Area, 0-1	0.63	0.2917	0.1838
Onsite Use of Resource, Hours per Year	72.23	0.0014	0.1011
Resident of Western Region of U.S., 0-1	0.22	0.3444	0.0758
City Size, over 2 million, 0-1	0.38	-0.1170	-0.0445
Education, Years	14.60	0.0248	0.3621
Household size, Persons	2.51	-0.1753	-0.4400
Age, years	46.13	0.0029	0.1338
Female respondent, 0-1	0.89	1.1181	0.9951
Black race, 0-1	0.04	1.3044	0.0522
Spanish heritage, 0-1	0.02	0.4727	0.0094
Combined question format, 0-1	0.41	0.0040	0.0016
Constant		2.8409	

[a] Adjusted R^2, 0.62; Cases, 134 participants; Dependent variable, logged number of occasions per year, mean, 63.25. Consumer Surplus, 1/0.0789 = \$12.68 per occasion of 1.76 hours. All variables are significant at the 0.01 level or better except household income, 0.88, and combined question format, 0.88.

Table 4. **Household Willingness to Pay Function for Protection of Valley Ranch Resources, United States, 1994[a]**

Variable	Mean	Coefficient	Elasticity
Valley Ranch Area Protection, Percent	89.45	0.0003	0.0268
Added Tax Referendum Vote, 0-1	0.80	0.2299	0.1839
Household Income Proxy, Hourly Wage	14.34	0.0358	0.5134
Leisure Time Resource, Days per Year	73.22	0.0012	0.0879
Site Quality Index, Miles	374.28	0.0005	0.1871
Interest in Direct Use, 1-3	2.12	-0.1314	-0.2786
Cost of Indirect Use per Occasion, Dollars	0.80	0.1101	0.0881
Importance of Indirect Use of an Option, 1-5	3.31	0.7352	2.4335
Expenses for Related Merchandise, Dollars per Year	7.12	-0.0171	-0.1218
Resident of Western Region of U.S., 0-1	0.14	1.1673	0.1634
Volunteer Time to Ranch Area Protection, Hours per Year	25.18	-0.0030	-0.0755
Combined Question Format, 0-1	0.47	-0.4516	-02122
Have Seen Ranch Areas, 0-1	0.47	-0.1708	-0.0803
Direct Use of Resource, Hours per Year	39.19	0.0025	0.0980
Indirect Use of Resource, Hours per Year	52.16	0.0008	0.0417
City Size, Over Two Million, 0-1	0.44	0.6588	0.2899
Age, Years	46.53	0.0098	0.4560
Female Respondent, 0-1	0.88	-0.3975	-0.3498
Spanish Heritage, 0-1	0.02	-0.1950	-0.0039
Education, Years	14.86	-0.0928	-1.3790
Household Size, Persons	2.44	0.0688	0.1679
Constant		1.0123	

[a] Adjusted R^2, 0.13; where the dependent variable is logged willingness to pay per year, with a mean of \$46.69 for 194 cases greater or equal to 1, representing 0.81 of households. All variables are significant at 0.01 level or better except Spanish Heritage, 0.05, and Ranch Area Protection, 0.26.

reading and watching valley ranch-related programs, as suggested by Bergstrom, et al. (1989; 1990); Whitehead and Bloomquist (1991), indicating the two activities are complements. WTP rises with the importance of option value, defined as having the opportunity (as a right) to visit valley ranch areas, i.e., the current enjoyment of anticipating the possibility of seeing valley ranch areas in the future even though they may not actually take future trips.

BENEFIT TRANSFER AND WELFARE DISTRIBUTION EFFECTS

The study was designed to explore the prospects for information transfer, that is, the benefit estimated for a specific site lacking data would be predicted by inserting appropriate values for explanatory variables into the statistical functions fitted to data from a national survey. The goal is to demonstrate how future inquiry into the subject of public benefit could contribute to the problem of estimating the total value of valley ranch open space in counties and states. This led us to examine the effects of the determinants of demand, and to look more closely at the details of the relative strength and positive or negative effect of the coefficients for explanatory variables than is usually the case in nonmarket valuation research. Most previous studies comparing expressed and revealed preference were primarily interested in the mean or median values obtained by variations in the two basic approaches to nonmarket valuation. While this study also is interested in the average values obtained by the two approaches, it differs in the level of interest in the shift variables.

Table 5 illustrates the net effect of adjusting three important variables for site specific conditions: region of the country, size of city, and distance from population centers. WTP and demand for direct and indirect use depend on many other variables, of course, which are held constant for purposes of illustration. The limited objective is to evaluate whether the net change

Table 5. Effect of Benefit Transfer Adjustments for Site Specific Region, Distance, and City Size of Respondents on Willingness to Pay and Demand for Direct and Indirect Use of Valley Ranch Resources, United States, 1994

Adjustment and Change in Value	Willingness to Pay	Demand for Trips to See Direct Use	Demand for Occasions of Indirect Use
Urban Fringe, 200 miles	-7%	7%	4%
Rural towns, suburbs, small cities	-66%	72%	12%
Region, West	117%	-30%	34%
Net Change	**44%**	**49%**	**50%**
Distance, 800 miles	28%	-29%	-17%
Rural towns, suburbs, small cities	-66%	72%	12%
Region, West	117%	-30%	34%
Net Change	**79%**	**13%**	**29%**
Distance, 800 miles	28%	-29%	-17%
City size, over 2 million	66%	-72%	-12%
Region, West	117%	-30%	34%
Net Change	**211%**	**-131%**	**5%**
Urban fringe, 200 miles	-7%	7%	4%
City size, over 2 million	66%	-72%	-12%
Regions, North and South	-117%	30%	-34%
Net Change	**-57%**	**-35%**	**-42%**
Urban fringe, 200 miles	-7%	7%	4%
Rural towns, suburbs, small cities	-66%	72%	12%
Regions, North and South	-117%	30%	-34%
Net Change	**-190%**	**109%**	**18%**

from the three adjustments in WTP are similar to the net change in demand for direct and indirect use. In three of the five trials, they are uniformly positive or negative, and in two of these, the net changes are of a similar magnitude. However, in two of the five trials, they are of different sign and very different amounts. This suggests that WTP studies may not be sufficient for benefit transfer purposes when there are direct and indirect uses of the resource.

Figure 2 summarizes statistical estimates of the effects of several demographic variables related to the difference between the distribution of WTP and consumer surplus from direct and indirect use. These include: income, education, region, distance from the resource, size of city, gender, household size, and race. Some demographic variables such as age, also are significant in the statistical equations for willingness to pay and consumer surplus from direct and indirect use. However, their effects are consistent, with the same sign across the three equations, and are not considered here. The coefficients of the explanatory variables in each of the three functions indicate the association between each variable and the dependent variable, given that all other variables included in the analysis are held constant. This means we are comparing groups of households that are similar with respect to every variable included in the function, except for the particular variable of interest. Elasticity shows the responsiveness of the dependent variable to changes in the price and nonprice variables included in the functions. The measure is particularly useful because it facilitates comparison among the variables in the three statistical functions independent of the units of measurement. Arc elasticities are shown for the relevant range of changes in the mean of variables.

In the WTP function, income is highly significant and positive. A one percent increase in the proxy for household income is associated with a 0.51 percent increase in WTP for preservation. Households with estimated annual income of $80,000 per year report they are

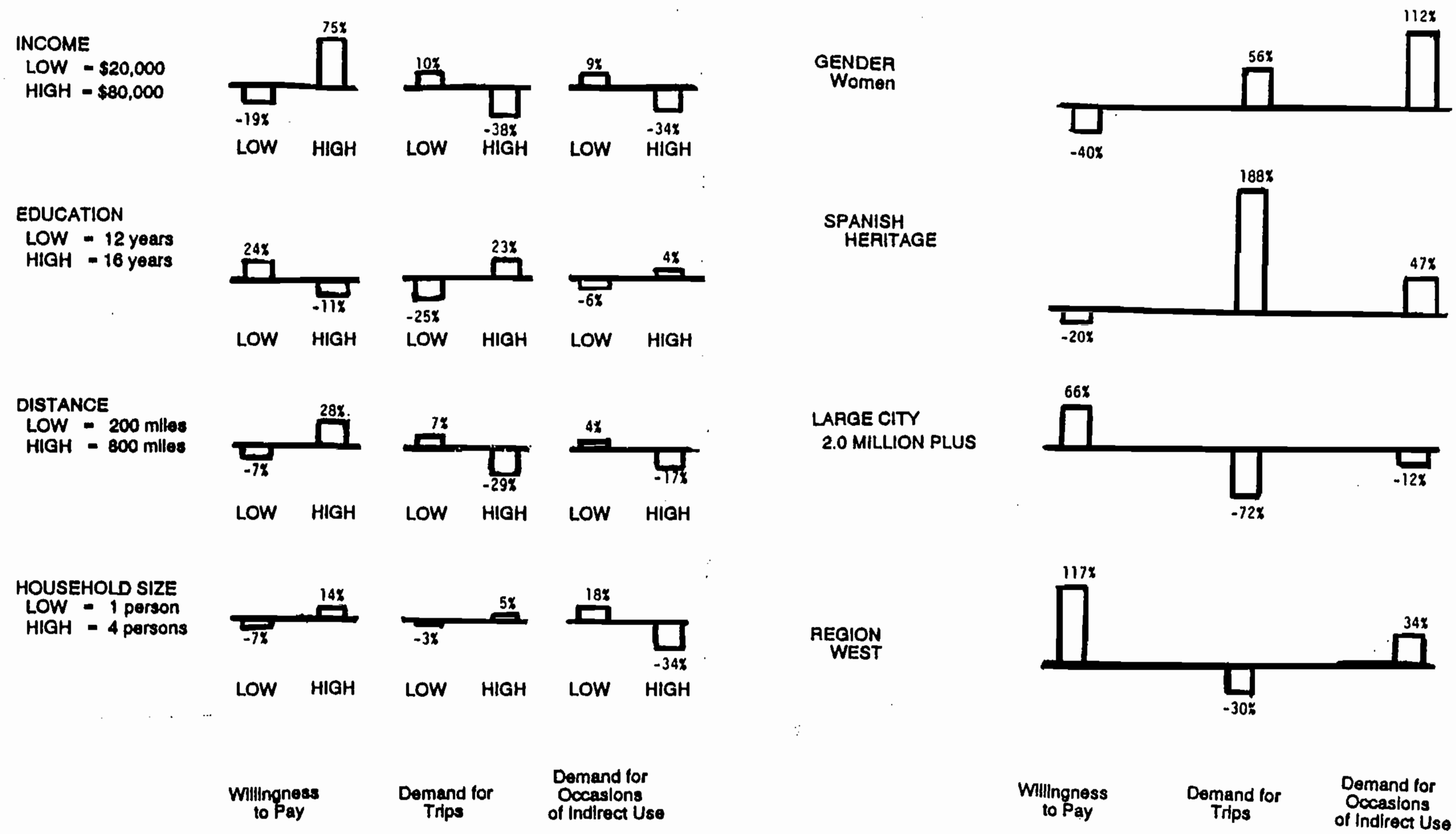


Figure 2. Distribution Effects of Household Income and Other Conditions on Willingness to Pay and Demand for Direct and Indirect Use of Valley Ranch Resources, United States, 1994

willing to pay 75 percent more than the national average WTP by middle income households with $40,000 annual income. Low income households with annual income of $20,000 per year are willing to pay 19 percent less than the national average, all else equal. The net effect is that high income households are WTP more than twice as much for preservation than low income households. While low income households report they are WTP less than high income households, their consumer surplus from direct and indirect use is over two-thirds greater. The average consumer surplus of direct and indirect use by households with annual income of $20,000 per year is estimated as $657 compared to $380 for households with income of $80,000. The net effect of low income is that while welfare measured by WTP is $36 less, welfare measured by consumer surplus of direct and indirect use is $277 more than for high income.

Household Income	Willingness to Pay	Consumer Surplus		Total
		Direct Use	Indirect Use	
$20,000	$31	$425	$232	$657
$80,000	$67	$239	$141	$380
Gain/Loss	-$36	$186	$91	$277

Low income households are able to participate more frequently in direct and indirect use by substituting greater inputs of own time for dollars per trip or occasion. Time has a much larger effect on direct and indirect use than money price. The two coefficients for the full price of direct and indirect use, money cost and opportunity cost of time, are both highly significant and negative. A one percent fall in money costs results in a 0.13 percent rise in occasions of indirect use and a 0.11 percent increase in trips for direct use. A one percent fall in opportunity cost of time results in a 0.23 percent rise in occasions of indirect use and a 0.26 percent increase in trips for direct use. Thus, given an equal percentage decrease in money and time cost for direct and indirect use, time has a 0.77-1.36 times larger effect on demand for direct and indirect use. Most households are in labor market disequilibrium where an hour of resource-related leisure time ($20) is perceived to be worth more than the average wage ($14). Also, these results are consistent with the price inelastic demand for most recreation activities, a one percent change in dollar cost or time cost results in less than a one percent change in quantity demanded (Walsh, 1986). This means that substitutes are usually not available to satisfy the perceived need, the proportion of income spent on it is low, and it is purchased frequently.

By way of comparison, earned income has a much larger effect than discretionary leisure time on WTP for preservation. Both the income constraint and time constraint are significant explanatory variables with the expected positive sign. Relaxing the income budget constraint by one percent, results in a 0.51 percent increase in WTP. Relaxing the time budget constraint variable, days of leisure time per year, by one percent, results in a 0.09 percent increase in WTP. Thus, given an equal percentage increase in the income and time budget constraints, WTP is 4.7 times more sensitive to income than available leisure time for this national sample of households. Thus, time appears to be a very weak substitute for income. A one percent increase in expressed willingness to volunteer time for preservation results in only a 0.08 percent decrease in WTP dollars.

These empirical results support the work by Hanemann (1991) that suggests the difference between reported WTP to avoid and WTA compensation for the loss of environmental services, may be due to the fact that direct and indirect use of public goods with few privately marketed substitutes significantly increases the total income (market plus in-kind services) of households, especially those with lower income. If this is so, it would seem to be important to get a better understanding of the sensitivity of this effect to the level of knowledge and productivity in nonmarket consumption.

In the WTP function, the explanatory variable, years of education, is highly significant and negative. With income constant, an increase in education of household heads is associated with a decrease in WTP. The arc elasticities show that in this case, college graduates are willing to pay 39 percent less than high school graduates. A review of the literature indicates these results are contrary to the findings of past studies which often report a significant positive effect of education on WTP for environmental preservation, which suggests that ranch resources and culture may be considered "low-brow." Another possible explanation may be that college graduates have become increasingly concerned about added government spending during the 1990s. With increased intelligence, college graduates may expect to obtain higher levels of resource preservation per added tax dollar than high school graduates.

The interesting result is that while college graduates are willing to pay less taxes for preservation, their consumer surplus from indirect use of the resource is 11 percent greater and from direct use 64 percent greater than high school graduates. This gap between nonmarket welfare effects of preservation for college and high school graduates is consistent with the trend toward higher benefits from knowledge evident in the labor market during the 1970s and 1980s. The results are contrary to the belief that while education and other investments in knowledge led to increased productivity and higher rates of return in the labor market, education has not had a similar effect on nonmarket productivity in the United States as compared to European education which emphasizes nonmarket consumption skills (Scitovsky, 1992). Also the relative preference by college and high school graduates suggests that it may not be "low brow" to be interested in direct and indirect use of ranch resources and culture (western clothing, music, art, fiction, history, etc.) in the 1990s.

Other WTP results in Figure 2 support the belief that households prefer preservation of remote high quality sites which primarily benefit the white male population living in large cities of more than 2 million in western states, and increases with size of household. However, the welfare distribution implications of these results would be misleading. The direct and indirect use results suggest that households prefer preservation of regional resources close to where they live even if lesser quality. These resources primarily benefit the nonwhite female population living in rural areas, suburbs, and smaller cities in the East and South, and decreases with size of household.

DISCUSSION

Several interpretations of the observed differences in stated and revealed preferences for preservation are possible. The most important is that respondents to a modified payment card WTP question do not appear to include the total utility and welfare impact of direct and indirect uses in reported added tax payment for preservation. Estimated average consumer surplus per household is several orders of magnitude greater than average WTP per household. The hypothesis of equality between benefit transfer of WTP values and direct and indirect use values is not supported in this case. This suggests that the NOAA panel recommended adjustment of WTP estimates may have the problem turned around. If the objective is to measure total utility and welfare impacts, then in addition to lowering WTP estimates by 50 percent for the effect of substitutes and budget constraints, they might consider possibly increasing benefit estimates by <u>several times</u> to compensate for loss of direct and indirect uses. It is not always correct to assume that WTP questions capture the total value of damages, including consumer surplus of direct and indirect uses.

Another interpretation is that CVM is essential to measuring existence value. If the objective is to estimate total utility and welfare impacts, it would not have been possible to measure existence value of the resource in this case by observed changes in behavior (spending money and/or time for direct and indirect uses). See the discussion in Larson (1993). To rely solely on measures of behavior would omit a substantial group of nonusers representing over 4 of 10 households that report positive WTP values for preservation. Only 46 percent of the households report direct and indirect use, including 43 percent indirect and 30 percent direct, while 81 percent report WTP a positive amount for preservation. Also, there is evidence from the positive coefficients for annual hours of direct and indirect use that revealed preference increases WTP by an estimated 13.4 percent (direct use, 9.3 percent; indirect use, 4.1 percent), so behavioral measures can contribute to understanding WTP, however, they cannot be considered a substitute for CVM surveys.

The most surprising result is that preservation of the resource increases the standard of living (broadly defined to include well-being) of the least advantaged people in society more than the most advantaged. Preservation primarily benefits the many low and lower middle income households who display a greater degree of resource dependency than higher income households. Thus, it is not correct to assume that WTP questions accurately measure the welfare distribution effects of preservation. This means that when the income variable has been insignificant as in several regional case studies of WTP for farm and ranch open space preservation (Bowker and Didychuk, 1994), it is not correct to assume that the utility and welfare impacts of preservation are distributed equally. Also, when the income variable has been significant and positive as in

Alaska state oil spill study (Carson, et al. 1992), it would not be correct to assume that the utility and welfare impacts of environmental preservation primarily benefit higher income households or the elite. Ranch open space is a form of natural capital that is non-substitutable and therefore the loss of which is irreversible. In cases where conversion of ranch land to urban subdivision occur, the unavoided foregone benefits are a social cost borne disproportionately by the poorest members of present and future generations, and they must be compensated for the loss of ranch open space services if Pareto efficiency is accepted as an objective of society.

Another possible interpretation of the results is to blame them on faulty theory or methodology. A different approach could reach the opposite conclusion. Before one can generalize these findings to other situations, replications with less familiar public goods than valley ranch resources are needed as well as testing with dichotomous choice questions and in-person interviews of a larger sample. Empirical testing of CVM has shown that dichotomous choice questions can produce WTP values several times higher than open-ended or modified payment card questions. CVM estimates from in-person interviews are usually larger than from mail surveys. The sum of monthly reported payment, as for the direct and indirect use questions, usually yield higher estimates of value than annual payment, used in the WTP question. WTP added taxes in the 1990s may be less than WTP small increases in the price of goods and services. The less familiar the public good, the lower would be the reported direct and indirect use. An unfamiliar site-specific public good may have no direct or indirect use value until information provided during a CVM interview stimulates immediate indirect use of knowing the resource exists and anticipation of possible future direct and indirect use.

Appropriate sized samples of 600 to 1,500 households would be more stable than small samples which have wider confidence intervals and are less suited for tests of significant difference. A TCM model that estimates the probability of participating would have a more elastic demand curve and lower consumer surplus than the model of number of trips per participant. We would expect a similar dampening effect on estimated demand for indirect uses. There is a need for research on the contribution of the resource to multiple purpose sightseeing trips, and the contribution of the resource to enjoyment of indirect use activities. Stated and revealed preferences represent somewhat different values, with WTP Hicksian compensating variation and direct and indirect use value Marshallian consumer surplus. Also, the context of stated and revealed preference questions may differ in important ways. The CVM questions about WTP place the respondent in the role of a citizen, while the direct and indirect use question context is that of producer and consumer of nonmarket activities.

CONCLUSIONS

The surprising results of this single study are not sufficient to conclude that it is necessary to include revealed preference questions in stated preference studies. However, the pilot study does emphasize the importance of empirically testing any recommendations that are made for improvements in CVM procedures. Randall and Stoll (1983) first argued more than a dozen years ago that household production of direct and indirect uses could influence CVM estimates of total value. Since the monetary measure of WTP represents behavioral intention rather than actual behavior, it is especially important to compare the results to measures of consumer surplus from direct and indirect use. It was feasible to introduce behavioral-based consumer surplus measures of value in this case since the resource was familiar to most respondents who had prior experience valuing it, and faced little uncertainty. The study demonstrates how future inquiry into the subject of direct and indirect use benefit could contribute to the problem of measuring the total value of valley ranch open space and other similar environmental resources. The results should be considered preliminary since there may be several possible influences or biases arising from the procedures applied in this case.

The results illustrate the possible welfare distributional effects when benefit measures rely on a single WTP question. Lower income households may have substantially less WTP values than higher income households. If resource allocation decisions are based on benefit-cost analysis applying similar WTP results, lower income households would be at a disadvantage, i.e., their collective preferences would be underrepresented. The resulting allocation decision could properly be called elitist and would be inequitable. However, a better representation of the collective preference of society would be based on some combination of household stated and revealed preferences, accounting for the higher income household ability to pay and the lower income household time resource. Resource allocations based on both of these benefit measurements would result in a more equitable distribution of the welfare effects of preservation.

To watch the details of what households and individuals do may be more important to understanding the distribution of welfare from resource preservation than what they say in response to a WTP question. The reason is that both time and income enter the nonmarket decision making process separately as constraints and prices of nonmarket activities under conditions of disequilibrium in labor markets (Bocksteal et al. 1987). Households pay both dollars and time. WTP dollars appears to understate utility of nonmarket resource-related activities in this case, which supports the work by Hanemann (1991) with respect to the divergence between WTP and WTA compensation. The larger consumer surplus of low income households is produced mostly by their own time input with relatively low opportunity cost compared to dollars, and WTA reflects the welfare effect of in-kind services of public goods.

REFERENCES

Adamowicz, Wiktor L., Jerald J. Fletcher, and Theodore Graham-Tomasi. 1989. "Functional Form and Statistical Properties of Welfare Measures." American Journal of Agricultural Economics 71:414-421.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Lerner, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." Federal Register 58(10):4602-4614. January 15.

Beasley, Steven D., William G. Workman, and Nancy A. Williams. 1986. "Estimating Amenity Values of Urban Fringe Farmland: A Contingent Valuation Approach." Growth and Change 17:70-78.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." Economic Journal 75:493-517.

Becker, Gary S. 1976. The Economic Approach to Human Behavior. University of Chicago Press, Chicago, Ill.

Bergstrom, John C., and H. Ken Cordell. 1991. "An Analysis of the Demand for and Value of Outdoor Recreation in the United States." Journal of Leisure Research 23(1):67-82.

Bergstrom, John C., B. Dillman, and John R. Stoll. 1985. "Public Environment Amenity Benefits of Private Land: The Case of Prime Agricultural Land." Southern Journal of Agricultural Economics 17:139-149.

Bergstrom, John C., John R. Stoll, and Alan Randall. 1989. "Information Effects in Contingent Markets." American Journal of Agricultural Economics 69:685-691.

Bergstrom, John C., John R. Stoll, and Alan Randall. 1990. "The Impact of Information on Environmental Commodity Valuation Decisions." American Journal of Agricultural Economics 70:614-621.

Bockstael, Nancy E., and Kenneth E. McConnell. 1981. "Theory and Estimation of the Household Production Function for Wildlife Recreation." Journal of Environmental Economics and Management 8:199-214.

Bockstael, Nancy E., Ivar E. Strand, and W. Michael Hanemann. 1987. "Time and the Recreational Demand Model." American Journal of Agricultural Economics 69:293-302.

Bowker, J. Michael, and D. D. Didychuk. 1994. "Estimation of the Nonmarket Benefits of Agricultural Land Retention in Eastern Canada." Agricultural and Resource Economics Review 23(2):218-225.

Cameron, A. Colin, and Pravin K. Trivedi. 1990. "Regression-based Tests for Overdispersion in the Poisson Model." Journal of Econometrics 46:347-364.

Carson, Richard T., Robert C. Mitchell, W. Michael Haneman, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1992. A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill. Report to the State of Alaska by Natural Resources Damage Assessment, Inc., La Jolla, CA. November. 127 p.

Cordell, H. Ken, and John C. Bergstrom. 1991. "A Methodology for Assessing National Outdoor Recreation Demand and Supply Trends." Leisure Sciences 13:1-20.

Creel, Michael D., and John B. Loomis. 1990. "Theoretical and Empirical Advantages of Truncated Count Data Estimates for Analysis of Deer Hunting in California." American Journal of Agricultural Economics 72:434-441.

Dillman, Don A. 1978. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley & Sons.

Dwyer, John F. 1994. Customer Diversity and the Future Demand for Outdoor Recreation. Gen. Tech. Rep. RM-252, Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture, Fort Collins, CO.

Foster, John H., John M. Halstead, and Thomas H. Stevens. 1982. Measuring the Nonmarket Value of Agricultural Land: A Case Study. Research Bulletin 672. Massachusetts Agricultural Experiment Station, University of Massachusetts.

Gardner, B. Delworth. 1977. "The Economics of Agricultural Land Preservation." American Journal of Agricultural Economics 59:1027-1036.

Halstead, John M. 1984. "Measuring the Nonmarket Value of Massachusetts Agricultural Land: A Case Study." Journal of the Northeastern Agricultural Economics Council 13:12-18.

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" American Economic Review 81:635-647.

Larson, Douglas M. 1993. "On Measuring Existence Value." Land Economics 69(4):377-388.

Loomis, John B., Armanado Gonzales-Caban, and Robin Gregory. 1994. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" Land Economics 70(4):499-506.

Lopez, Rigoberto A., Farhed A. Shah, and Marilyn A. Altobello. 1994. "Amenity Benefits and the Optimal Allocation of Land." Land Economics 70(1):53-62.

McConnell, Kenneth E. 1989. "The Optimal Quantity of Land in Agriculture." Northeastern Journal of Agricultural and Resource Economics 18:63-72.

McConnell, Kenneth E. 1992. "On-Site Time in the Demand for Recreation." American Journal of Agricultural Economics 74:918-925.

McKean, John R., Donn M. Johnson, and Richard G. Walsh. 1995. "Valuing Time in Travel Cost Demand Analysis: An Empirical Investigation." Land Economics 71(1):96-105.

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

Randall, Alan, John P. Hoehn, and Cindy S. Swanson. 1990. Estimating the Recreational, Visual, Habitat, and Quality of Life Benefits of Tongass National Forest. Gen. Tech. Rpt. RM-192. Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture, Fort Collins, CO.

Randall, Alan, and John R. Stoll. 1983. "Existence Value in a Total Valuation Framework." in Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Robert D. Rowe and Lauraine G. Chestnut, eds. Westview Press, Inc., Boulder, CO.

Robinson, John P. 1991. Free Time in America. University of Maryland, College Park. Draft.

Rosenberger, Randall S., Richard G. Walsh, John R. McKean, and C. J. Mucklow. 1995. Benefits of Ranch Open Space to Local Residents. Bulletin XCM-____, Colorado State University Cooperative Extension, Fort Collins. Forthcoming.

Rosenthal, Donald H. 1987. "The Necessity for Substitute Prices in Recreation Demand Analyses." American Journal of Agricultural Economics 69(4):828-837.

Scitovsky, Tibor. 1992. The Joyless Economy: The Psychology of Human Satisfaction. Revised Edition. Oxford University Press, New York.

Smith, V. Kerry, and Yoshiaki Kaoru. 1990. "Signal or Noise? Explaining the Variation in Recreation Benefit Estimates." American Journal of Agricultural Economics 72:419-433.

U.S. Department of Agriculture. 1985. Farm Real Estate: Historical Series Data: 1950-1984. Statistical Bulletin 738. Washington, DC.

U.S. Water Resources Council, 1983. Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies. U.S. Government Printing Office, Washington, DC.

Walsh, Richard G. 1986. Recreation Economic Decisions: Comparing Benefits and Costs. Venture Publishing, Inc., State College, PA.

Walsh, Richard G., Donn M. Johnson, and John R. McKean. 1990. "Nonmarket Values from Two Decades of Research on Recreation Demand," in Advances in Applied Micro-Economics. V. Kerry Smith and Albert N. Link (eds.). Vol. 5, JAI Press, Greenwich, CT. p. 167-194.

Walsh, Richard G., and John R. McKean. 1992. "Indirect Option Value," in Benefits and Costs in Natural Resource Planning, R. Bruce Rettig (ed.), USDA Western Regional Committee W-133, Oregon State University, Corvallis.

Walsh, Richard G., John R. McKean, and Randall S. Rosenberger. 1995. National Household Benefits of Preserving Valley Ranch Open Space. Bulletin XCM-______, Cooperative Extension, Colorado State University, Fort Collins. Forthcoming.

Walsh, Richard G., John R. McKean, Randall S. Rosenberger, and C. J. Mucklow. 1994. Recreation Value of Ranch Open Space. Bulletin XCM-167, Cooperative Extension, Colorado State University, Fort Collins.

Walsh, Richard G., Larry D. Sanders, and John R. McKean. 1990. "The Consumptive Value of Travel Time on Recreation Trips." Journal of Travel Research 29(1):17-24.

Whitehead, John C., and Glenn C. Blomquist. 1991. "A Link Between Behavior, Information, and Existence Values." Leisure Sciences 13:97-109.

Willis, Ken G., and M. Charles Whitby. 1985. "The Value of Green Belt Land." Journal of Rural Studies 1:147-162.

Young, Trevor, and P. Geoffrey Allen. 1986. "Methods for Valuing Countryside Amenity: An Overview." Journal of Agricultural Economics 37:349-361.

WTP Certainty Intervals and the Disparity between Contingent Valuation Elicitation Formats: Evidence from an Existence Values Experiment

Gregory L. Poe and Michael P. Welsh*

Abstract: This paper presents the results of a study of differences in existence value estimates associated with alternative contingent valuation elicitation formats. Modifying a pilot survey instrument designed for valuing reduced fluctuations in Glen Canyon dam releases, a split sample design was conducted on 626 Cornell students in four classes. Approximately one-third of the students completed a standard dichotomous choice questionnaire, one-third completed a payment card version, and one-third completed a multiple bounded discrete choice format in which respondents are allowed to express their level of willingness to pay (WTP) certainty for different dollar thresholds. All three response formats were analyzed using a switching or bounded distribution likelihood function.

The ratio of dichotomous choice to payment card mean WTP value estimates was 2.7:1. Comparisons of the multiple bounded discrete choice format with the dichotomous choice and payment card responses indicate that the multiple bounded discrete choice format cover the range of values associated with the other two elicitation methods. Moreover, alternative parameterizations of the multiple bounded discrete choice model correspond with the dichotomous choice and payment card response functions. One interpretation of these results is that the payment card and dichotomous choice elicitation measures invoke different weighting schemes of uncertain preferences.

* Assistant Professor, Department of Agricultural, Resource, and Managerial Economics, Cornell University, and Senior Associate, Hagler Bailly Consulting Incorporated. Senior authorship is shared. The authors are indebted to Dick Aplin, Cindy van Es, Ed McLaughlin, Bill Schulze and their classes for participating in this experiment. Rich Bishop, Kevin Boyle, Rich Ready, Bill Schulze, and W-133 participants offered helpful insights on this project. The normal disclaimers apply. Funding was provided by Cornell University Agricultural Experiment Station under Regional Project W-133 Benefits and Costs Transfers in Natural Resources Planning. Permission to use and modify pilot survey instruments developed by Glen Canyon Environmental Studies (U.S. Bureau of Reclamation) is greatly appreciated.

WTP Certainty Intervals and the Disparity between Contingent Valuation Elicitation Formats: Evidence from an Existence Values Experiment

1. Introduction

Contingent valuation practitioners, experimental economists, and psychologists have long recognized that the use of different elicitation formats can result in divergent value estimates [Cummings *et al.*; Grether and Plott; Slovic and Lichtenstein]. Recently, however, the importance of elicitation effects has been elevated in the controversy over the use of continent valuation to measure passive use, or existence, values. At one extreme the NOAA panel focused attention on elicitation mechanisms by identifying the dichotomous choice technique as the preferred, conservative value elicitation method [Arrow et al.]. At the other extreme, critics of contingent valuation have suggested that violations of "procedural invariance" or failure to demonstrate "convergent validity" across value elicitation formats form a basis for rejecting the validity of contingent valuation altogether [e.g., McFadden; Desvousges *et al.*; Schkade and Payne].

We believe that it is premature to dictate a preferred valuation format and that differences in values do not necessarily indicate that the technique is unreliable. Both issues merit further investigation in controlled experiments and actual valuation studies in order to better quantify or calibrate differences between elicitation formats, and to begin to understand factors that contribute to these differences.

In this paper we provide a direct comparison of the dichotomous choice and payment card elicitation techniques, two widely used contingent valuation methods that, to our knowledge, have not been directly compared in previous research[1]. Our results support other research that dichotomous choice values do not provide conservative estimates [see Ballestreri *et al.* for a review of these studies]. This paper also provides the first empirical comparison of a multiple bounded discrete choice (MBDC) elicitation format with the dichotomous choice and the payment card techniques. The MBDC allows respondents to express their level of voting certainty (ranging from "definitely no" to "definitely yes") for a range of referendum thresholds. An intriguing result of this experiment is that the willingness to pay distribution estimated from the payment card data corresponds to the "probably yes" MBDC response patterns, while the dichotomous choice responses corresponded with the "unsure" MBDC response patterns. In this manner, uncertainty over preferences may be a contributing factor in observed differences across elicitation formats.

The remainder of this paper is organized as follows. The second section proposes that variations in values across elicitation formats can be attributed to uncertainty in preferences and elicitation specific decision heuristics. The third section develops a MBDC method that allows individuals to specify their level of decision certainty. Data collection procedures and results are provided in the fourth section.

---

[1] Boyle and Bishop found that mean WTP estimates from income anchored payment cards exceeded dichotomous choice values by a factor of 1.6:1, but this payment card technique has never been widely used and is likely to be subject to anchoring effects. Cameron and Huppert found the dichotomous choice values exceeded payment card values in monte carlo simulations from the same underlying WTP distribution. While monte carlo simulations do provide interesting insights, the use of simulations abstracts away from any contextual effects associated with different elicitation formats. In making this statement we rely on previous research in experimental economics that demonstrates that theoretically isomorphic auction mechanisms provide different results [Davis and Holt; Kegel and Roth], or psychology which argues simply that decisions will be contextual [Payne *et al.*].

2. Preference Uncertainty and Willingness to Pay

To paraphrase Kenneth Arrow [1986] and others, "If you ask a contingent valuation question, you will get an answer". Economic analyses based on axiomatic models of preferences and rational choice, contend that presumably equivalent elicitation procedures should demonstrate procedural invariance. That is, individuals assumed to have well ordered preferences should provide similar and consistent answers whether they are asked to assign a value to a good, as in the open ended or payment card elicitation methods, or asked to make a direct choice, as in the dichotomous choice or referendum technique.

Twenty five years of research in behavioral decision making consistently demonstrates that the assumptions about preferences and the economic theoretic requirements of procedural invariance are not frequently observed in actual choices. The majority of these experiments have been limited to simple gambles and experimental choices. However, a study by Magat *et al.* of people's values for chemical product risks demonstrates that invariance does occur in comparisons of contingent ranking and contingent valuation. Similarly, Irwin *et al.* demonstrate preference reversals in willingness to pay for air quality improvements and choice between alternate improvement levels and private goods.

Although procedural invariance might be expected to hold when individuals have "well articulated preferences and beliefs", this body of research demonstrates that individuals simply do not have stable, exogenous, and consistent values and beliefs [Slovic and Lichtenstein; March; Tversky *et al.* 1988, 1990]. The practice of rational choice is not simple, and necessarily involves two types of guesses: "guesses about future consequences of current actions and guesses about future preferences for those consequences" [March, p. 589]. Within this framework of preference uncertainty, it is argued that different value elicitation frameworks invoke different decision heuristics, resulting in seemingly inconsistent values. In a conclusion from this research that is particularly relevant to the current study, Tversky *et al.,*1988, note:

> Real world decisions can sometimes be framed either as a direct choice (e.g., Should I buy the used car at this price?) or as a pricing decision (What is the most I would pay for that used car?). Our findings suggest that the answers to the two questions are likely to diverge [p. 383]

As such, even simple, familiar "market" decisions are <u>expected</u> to result in procedural invariance across elicitation formats. Because individuals rely more heavily on simplifying heuristics as the context of the decision becomes more complex, it is likely that elicitation effects will increase when valuation decisions involve environmental or other "exotic" goods characterized by ambiguity or uncertainty in preferences, ambivalence or conflicting values, and non-comparable trade-offs [Payne *et al.*; Schkade and Payne].

In a series of papers, Opaluch, Opaluch and Segerson, and Ready *et al.* have incorporated this notion of preference uncertainty or ambivalence into the standard economic framework of indifference curves and preference mapping. These analyses explicitly recognize that environmental commodities involve often difficult tradeoffs between tastes, morals, and environmental values. Moreover, the process of assigning monetary values to environmental commodities is a novel exercise for most individuals, and they often lack time to fully introspect their preferences when making decisions. Following the arguments in Ready *et al.*, the consequences of uncertainty over preferences is depicted in Figure 1, which considers reference and alternative options in money/amenity space. Point I represents the initial or reference levels of money and amenities. Points to the Northeast of I represent possible alternative combinations, which, if preferences are monotonically increasing in money and amenities, will be strictly superior to I. Conversely, the southwest quadrant depicts money/amenity pairs that are strictly inferior. In either of these cases, no trade-offs are required, and preference between the reference conditions and alternatives in these two quadrants are obvious.

Yet, by design, contingent valuation studies are concerned with trade-offs between environmental amenities and money, measuring either willingness to pay (WTP) for environmental improvements, or willingness to accept (WTA) compensation to experience environmental degradation. Focusing on WTP measures depicted in the northwest quadrant[2], individuals might have no difficulty in making choices in obviously imbalanced options. Using a well-known example, most individuals would probably not spend a large amount of money to save a few birds from dying in waste-oil holding ponds. In contrast, individuals might be willing to spend a nominal amount to save a large portion of a threatened bird population. However, in cases where there is preference uncertainty or ambiguity, there will be a range of options in which choices between money and environmental amenities are not so clear cut and the respondent will have difficultly in evaluating the trade-off between money and the amenity. This so-called ambivalence region [Ready *et al.*] is depicted by the area bounded by V_1IV_2.

The existence of an ambivalence region translates into uncertainty in WTP in

[2] This is not intended to discount WTA as a valid welfare measure, but the analysis is parallel and thus unnecessary at this juncture. Uncertainty in WTP and WTA have been used to explain WTP/WTA disparities [Hoehn and Randall; Opulach and Segerson; and DuBourg *et al.*].

contingent valuation surveys. In terms of Figure 1, the uncertain individual would definitely pay W_L or less for a ΔA change in the level of amenity, and definitely not pay W_U or more. That is, even when uncertainty in preferences exists, there are values which the individual would definitely pay or definitely not pay, corresponding to money/amenity combinations outside the ambivalence region. Yet, for dollar values falling in the range bounded by W_L and W_U , there will be a degree of preference uncertainty. In probabalistic terms there is a probability of one that the individual would pay a posted offer of $\$<W_L$ and reject any posted offer in which $\$>W_U$. As such, it is possible to envision a cumulative WTP distribution W(A) that approaches zero at W_L and 1 at W_U[3].

Conventional economic approaches underlying economic conceptualizations of WTP [e.g., Freeman] assume away this uncertainty in preferences. In this world of well-ordered and well-articulated preferences, a single indifference curve is identified, setting $W_L=W_U$. However, recent evidence from contingent valuation research suggests that this value uncertainty exists and may be substantial. In the Alaska study of the Exxon Valdez, Carson *et al.* provide indirect evidence of uncertainty by including unsure options in dichotomous choice formats and assessing the strength of convictions of the 'yes' respondents: almost 10 percent of the respondents were unsure if they would vote yes or no for the program at the specified dollar amount; of the respondents who voted yes, 23 percent did not strongly favor the program at that price. A study by Li and Mattsson followed up the dichotomous choice

[3] The ambivalence region might be expanded to include uncertainty over provision levels, as suggested in Hoehn and Randall and evaluated especially in Svento. In the context of Figure 1, this would add a vertical uncertainty band associated with perceived ΔA outcomes, thus widening the $W_L W_U$ range.

question with a confidence question, "How certain were you to your answer in the previous question?" Only 26 percent indicated that they were 100 percent certain of their yes or no response. Finally, DuBourg *et al.* explicitly quantified imprecision or uncertainty in WTP for reduced risks, identifying a "best point estimate" of WTP (W_B) along with the "smallest amount that a respondent would *not* pay" (W_u) and the "largest amount s/he would definitely pay" (W_L). The conclusion that "people did not feel definite" (p. 129) about their preferences was supported by wide deviations between W_L and W_U.

The relationship between this observed uncertainty and procedural invariance across elicitation formats might be attributed to alternative heuristics used to report single values or choices. In reporting values, different heuristics might be employed, including lexicographic rankings, minimax or maximin strategies, safety fixed criteria, or various forms of weighting schemes. To the extent that different elicitation methods invoke different decision processes, say $G_1(W(A))$ and $G_2(W(A))$, we would expect different results. For example, the conceptual analysis by Hoehn and Randall and the empirical results by DuBourg *et al.* suggest that open-ended WTP questions elicit responses closer to the lower end of the WTP uncertainty interval. In contrast, observed disparities between dichotomous choice and true market values [Ballestreri *et al.*] and evidence of overstatement in posted offer markets [Davis and Holt; Kagel and Roth] and yea-saying [Kanninan] suggest that dichotomous choice markets invoke a decision heuristic that tends to overstate "true" WTP.

The suggestion that differences in contingent value across elicitation methods can be attributed to different decision heuristics applied to a common WTP distribution differs from the psychological notion that contingent values derive from constructed rather than revealed preferences [Schkade and Payne; Payne *et al.*; Gregory *et al.*; Tversky *et al.*, 1988]. Here, it is assumed that individuals have underlying preferences for exotic as well as familiar commodities, but that there is uncertainty or a fuzziness in these preference functions due to limited information, experience, or interests. Rather than creating or constructing their preferences from broad basic values, they instead adopt weighting schemes across prior preference relationships. Some empirical support for this interpretation is found in a study by Kealy and Turner which compared within subject dichotomous choice and open-ended valuation responses. In spite of the result that dichotomous choice values for a public good were 1.4 to 2.5 times as large as open-ended values, there were high correlation coefficients, ranging from 0.53 to 0.95, between responses in jointly estimated models. Thus, although individuals responded differently according to question type, responses were not independent. The approach suggested here is also consistent with the microeconomic system framework of experimental economics [Smith, 1982, 1987], in which respondent-specific preferences, household technology, and attitudes and preferences are brought to bear on changes in quantities and qualities in a created institutional framework [Coursey and Schulze; Bergstrom and Stoll]. Rather than arbitrarily shifting values as suggested by a constructed preferences orientation, the microeconomic systems framework asserts that respondents' reported values are a function of prior individual preferences. Institutional and information effects are expected, but they are tempered by existing preference functions.

3. Multiple Bounded Discrete Choice (MBDC) Format

Sociologists have developed what has come to be known as the "return potential" question format [Jackson; Shelby] to explore the strength of social norms (e.g., crowding) and satisfaction levels across varying conditions (e.g., boating parties encountered). The typical question format associated with this normative approach is a two dimensional matrix, in which one dimension delineates differing levels of the commodity and the other elicits some ordered response category. For example, in a study of the effects of streamflow on Grand Canyon river rafting recreational benefits, Shelby *et al.* asked respondents to rate 14 different streamflow levels ranging from 1 (very satisfactory) to 5 (very unsatisfactory).

In a recent paper, Welsh and Bishop demonstrated that the (ordinal) return potential format could be adapted to provide (cardinal) contingent valuation estimates of WTP by describing a referendum over whether or not to pursue public provision of a non-market good. The dimension over which respondents are asked to make choices is the dollar amount that they would be required to pay if the referendum passed. The other dimension allows individuals to express their certainty levels or intensity of preferences for the referendum at each dollar value. An example of this approach is provided in Figure 2.

This proposed MBDC approach contains elements of, and builds upon, both the payment card and dichotomous choice approaches widely used by contingent valuation practitioners. Like the payment card format, respondents are presented with an ordered sequence of thresholds. Yet, rather than circling a single value or interval, the respondent is given a "polychotomous choice" response option including, say, "definitely no", "probably no", "unsure", "probably yes", and "definitely yes", that allow an expression of WTP certainty for different dollar thresholds[4]. In this manner, the context of the good to cost tradeoff is expanded beyond traditional dichotomous choice questions to include additional dollar options and strength of conviction. For the researcher, the multiple bounded approach is more efficient than single-bounded or double-bounded choice models, thus reducing sample size requirements [Welsh and Bishop]. Complexities associated with optimal bid designs in choice based models are similarly avoided [e.g., Kanninen, 1993; Cooper and Loomis].

Analysis of maximum WTP associated with each certainty level (WTP_C) is conducted using a multiple bounded generalization of double bounded models [Hanemann *et al.*; Welsh and Bishop], analogous to the maximum likelihood interval modeling approach used for payment card data [Cameron and Huppert]. Defining X_{iL} as the maximum posted price that the individual would vote for in the response category or higher, and X_{iU} to be the lowest posted price that the individual would not vote in the response category, WTP_{iC} lies somewhere in the switching interval $[X_{iL}, X_{iU}]$. Letting $F_C(X;\beta)$ denote a statistical distribution function with parameter vector β, the probability that an individual would respond with lower than certainty level C to a posted threshold would be $F_C(X;\beta)$. The associated probability of a response level greater than or equal to C to a given amount X would be given by $1\text{-}F_C(X;\beta)$. In all, the contribution to the joint likelihood function across all individuals for this response category would be $F_C(X_{iU};\beta) - F_C(X_{iL};\beta)$. The corresponding log-likelihood function is given by

[4] Recent articles by Svento, and Ready *et al.* use polychotomous choice responses to single referendum questions.

$$\ln(L) = \sum_{i=1}^{n} \ln[F_C(X_{iU};\beta) - F_C(X_{iL};\beta)] , \tag{1}$$

which is essentially a multiple bounded equivalent of the double bounded models employing dummy variables for each response category[5] [Hanemann *et al.*; Welsh and Bishop].

Using the bounded switching model approach, three different MBDC response functions are reported in the following section. The "definitely yes" model corresponds to modeling the interval at which individuals switch from "definitely yes" responses to a lower certainty level. Similarly, the "probably yes" model is based on respondent-specific switching intervals from the "probably yes" response to lower certainty levels. Finally, the "unsure" MBDC model estimates the switching range from unsure to "probably no" or "definitely no" responses. In this manner, the unsure MBDC model does not refer to a situation explored in other surveys [e.g., Carson *et al.*], in which the respondent expresses ex post uncertainty in his or her responses. Rather, the "unsure" MBDC model estimates the point at which respondents switch from extreme preference uncertainty (or indifference) to not preferring the program at the posted price. These switching points are readily identified using a simple computer program.

Dichotomous choice and payment card response models can be treated as special cases of this bounded distribution likelihood function. Extension of this analysis to payment card

---

[5] This modeling approach assumes that individuals resolve their uncertainty in choosing the polychotomous answer rather than directly incorporating uncertainty into the estimation process. Svento offers an alternative approach incorporating uncertainty over outcomes into the estimator, thus retaining the assumption that "the utility function is nonstochastic (no random variation in tastes)" [Hanemann, 1985, p. 10]. Li and Mattsson introduce an alternative approach in which there is a composite error function arising from omitted variables and from preference uncertainty.

data simply involves redefining $F_C(.)$ to be the probability that WTP<X, and, thus, evaluating only ranges bounding maximum WTP [see Cameron and Huppert]. With the same redefinition of $F_C(.)$, "yes" responses to dichotomous choice data can be incorporated into the log-likelihood function by setting $F_C(X_{iU};\beta)=1$. For "no" responses to dichotomous choice thresholds, $F_C(X_{iL};\beta)$ is set to zero [see Hanemann, 1984].

The hypothesis of equality between estimated maximum likelihood response function can be evaluated using the likelihood ratio test:

$$\mathrm{LR} = 2*[(ll_1+ll_2)-ll_{pool}] \sim \chi^2(r) \tag{2}$$

where ll_1 and ll_2 are associated with the individual models to be compared, ll_{pool} is the log likelihood value associated with the pooled model that imposes the equality of coefficients (β), and r is the number of restrictions. A difficulty associated with performing this test is that estimation of the pooled model requires a similar data structure across individual models. For comparison across the alternative MBDC switching functions this requirement obviously poses no problem. Similarly, pooling the MBDC switching function with the PC bounded function is straightforward. However, it is not immediately obvious how these formats could be compared with DC responses, in which only a yes or no response is obtained from a single bid amount. This is accomplished by creating an artificial MBDC data set for DC data. Note that if the MBDC response exceeds the highest threshold value (i.e. yes to all the n possible values), $F_c(X_{iu};\beta)$ is set equal to 1. Alternatively, if the MBDC response is below the lowest value (i.e. no to all n values) then $F_c(X_{iL};\beta)$ is set to zero. This relation can be exploited in converting a DC response to a multiple bounded response. If a respondent answers yes, then n-1 artificial threshold are created below the DC bid and $F_c(X_{iu};\beta)$ is consequently set equal to 1. If instead the response is no to the DC bid, then n-1 artificial thresholds are created above the DC bid. In this manner, the data sets are of similar construction and can be estimated jointly.

IV. Data Collection and Results

The primary purpose of this study was to further explore the MBDC method and its relationship to the dichotomous choice (DC) and payment card (PC) elicitation formats. Previously, the MBDC method had only been evaluated statistically using simulation techniques from known distributions [Welsh and Bishop], and field application in small focus groups or survey pretests. Given the similarities to DC and PC, there was also a desire to investigate the correspondence, if any, between the three methods.

On this basis, a split sample experiment was designed, in which approximately one-third of the sample received a DC question, one-third completed a PC version, and one-third completed a MBDC question. Rather than design a completely new survey instrument, a pilot study of WTP for reduced fluctuations in Glen Canyon Dam releases, developed by Glen Canyon Environmental Studies, was used. This survey instrument had been extensively pre-tested and approved for a federally funded contingent valuation survey. Importantly, the scenario elicited an existence value based WTP for protecting Grand Canyon beaches, and associated ecological and native American archeological preservation values. It is precisely these conditions that one would expect the greatest uncertainty in preferences and values.

The Glen Canyon Environmental Studies pilot survey was modified in length and content to allow for classroom distribution. The resulting nine page questionnaire and associated six page information sheet was distributed to four Cornell undergraduate classes in the Department of Agricultural, Resource, and Managerial Economics. In each class, the objectives of the experiment were discussed, and briefly related to the subject matter in the course. An incentive of a $100 public lottery to be drawn from returned, fully completed questionnaires was provided for each class. Two classes allocated 30 minutes for completing the questionnaire in the classroom, with an associated response rate of 96 percent (498/521). The students in the other two classes were instructed to return their completed questionnaires at the following class meeting, providing a lower response rate of 48 percent (128/226).

Before completing the questionnaire, respondents were instructed to first read a six page information sheet consisting of a map and information about the dam, the study area, and the relationship between the Glen Canyon Dam and the study area. Particular emphasis was given to describing the existing natural resources, archeological sites, and fish populations, and the present and future impact on these resources as a result of changes in instream flows associated with the operation of the hydroelectric facility at the Glen Canyon Dam. The existence value component was emphasized in the information sheet by noting that only a small percentage of the visitors to the Grand Canyon National Park actually see or use the resources in the study area, and that "the only people who see the resources in the study area are American Indians using resources in the study area, people who raft and/or backpack, and people who fish there."

A comprehensive true/false quiz at the beginning of the questionnaire encouraged reading and assimilation of the information provided. Following this quiz, a hydroelectric management "proposal" (Figure 3) that would eliminate daily fluctuations in the river level and mimic, to the extent possible, natural seasonal fluctuations was described. In each of the different elicitation formats, respondent willingness to vote for the proposal if it did not personally cost them anything was elicited. Participants who indicated that they would vote for the proposal at zero personal costs were then asked one of the three contingent valuation question depicted in Figures 2, 4 and 5. The MBDC (Figure 2) and PC (Figure 4) questions consisted of identical series of 13 dollar thresholds ranging from 10¢ to $200 per annum. DC bids were individually inscribed in the blank space in Figure 5, with each DC participant receiving one of nine values between $1 and $200 ($1, $5, $10, $20, $30, $50, $100, $150, $200).

Voting patterns on the "no cost" proposal in Question 2 (Figure 3) were similar across elicitation formats, as would be expected by the fact that this question preceded the contingent valuation questions: 93.2, 95.5, and 93.2 percent of the MBDC, PC, and DC respondents voted yes to the "no cost" proposal. Conditional response patterns for the "yes" respondents to the "no cost" proposal are provided in Table 1. The first column indicates the range of posted dollar thresholds. The next five columns provide the distribution of MBDC responses for each posted threshold. PC and DC responses are indicated in the last two columns.

PC and DC responses follow expected patterns. PC responses reach a mode at $10 and are positively skewed. The proportion of DC "yes" responses generally declines as threshold values increase. Within these general trends, response spikes are found at important rounded thresholds (e.g., $10 and $100). Similarly, the distribution of MBDC responses follow expected patterns. The "definitely yes" responses drop and the "definitely no" responses rise monotonically with threshold values. "Probably yes", "unsure", and "probably no" responses peak at intermediate values, with modes at $30, $75, and $150, respectively.

Estimation of PC, DC response functions and the "definitely yes", "probably yes" and "unsure" MBDC switching function used the bounded likelihood function in Equation 1, and the standard logistic function for the cumulative distribution function.

$$F_c(X;\beta) = \frac{1}{1+e^{-(\alpha+\beta X)}} \tag{3}$$

As shown in Table 2, the estimated constant and slope coefficients in each of the five models were significant at the 5 percent level. The corresponding logit models are graphically depicted in Figure 6. As expected, the distribution of MBDC responses is monotonic, with the cumulative distribution function shifting right with a decline in positive response certainty. The graphical depiction also indicates that there is a close correspondence between the DC and the "unsure" MBDC estimates, and the PC and "probably yes" MBDC model.

Statistical equivalence of the different logit models was evaluated with the following hypothesis test:

$$H_o^1:\ F_1(X;\beta) = F_2(X;\beta)$$

Of the 10 possible pairwise hypothesis tests only $F_{PC}(.) = F_{\text{Prob. Yes, MBDC}}(.)$ and $F_{DC}(.) = F_{\text{Unsure, MBDC}}(.)$ were not rejected at the 5 percent level using the likeihood ratio test in equation (2). All other comparisons were rejected. These results suggest that individuals who are unsure about their WTP as indicated by willingness to pay responses to a specified dollar amount, would respond "yes" to a similar DC threshold value. Such an interpretation conflicts with the Ready *et al.* polychotomous choice analysis that DC respondents switch from "yes" to "no" responses close to the lower bound of the ambivalence region[6], but is consistent with past research indicating yea-saying and overstatement in DC and other posted offer markets [Kanninen, 1995]. In contrast, the estimated PC response function exhibits a pattern similar to that of the "probably yes" MBDC model, implying that PC respondents are fairly certain about their WTP statements. This latter result is consistent with the DuBourg *et al.* finding that continuous WTP responses were fairly conservative, and tended towards the lower bound of the uncertainty interval.

Analytical mean ($-\frac{1}{\beta}\ln(1+e^{\alpha})$) and median ($-\frac{\alpha}{\beta}$) values of WTP were calculated from the parameter estimates, and 95 percent confidence bounds were estimated using the Krinsky and Robb parametric bootstrap procedure [Park *et al.*]. These values are depicted in Table 3. Consistent with the results of H_0^1, there is a close correspondence between the mean and median distributions of PC and the "probably sure" MBDC models and the DC and "unsure", MBDC models. Difference of means and median tests were conducted for each of the 10 comparisons using the convolutions approach developed in Poe *et al.*. Hypotheses of equality

---

[6] The Ready *et al.* results may be attributed to the arbitrary definition of ambivalence zone bounds. Their polychotomous choice format included six response option related to the direction and strength of preferences between the reference and target condition, but did not include an unsure response option. Instead, the lower bound of the ambivalence zone was defined as the lowest dollar amount to which 50 percent of the respondents would respond "probably no" or "definitely no". The upper bound was the highest dollar amount to which 50 percent would give a "probably yes" or a "definitely yes" response. The middle of the ambivalence region was defined as the dollar value at which 50 percent of the respondents would say "maybe yes" or higher. In three studies investigated, dichotomous choice responses fell below the middle of the ambivalence region.

were not rejected for the PC/"probably yes", MBDC and the DC/"unsure" MBDC mean and median comparisons. All other pairwise comparisons were rejected at the 5 percent level.

Importantly, the disparity between DC and PC formats is significant and large. Estimated 95 percent confidence bounds do not overlap, and DC/PC ratios for the mean and median are 2.7:1 and 2.9:1, respectively. These ratios correspond with the range of values from contingent valuation studies comparing open-ended and DC responses.

V. Conclusions and Implications

Our results indicate that the MBDC format provides a range of values encompassing independent DC and PC values. As such, this format offers the potential for an efficient means of estimating a range of WTP estimates associated with two widely used elicitation formats.

The observed 2.7:1 mean WTP ratio between DC and PC responses appears to be associated with different cognitive processes in resolving preference uncertainty. Using an existence valuation scenario in which respondents are likely to have substantial uncertainty in preferences, DC responses correspond with MBDC models in which respondents indicated that they were "unsure" whether they would vote for the proposal at a specified threshold price. In contrast, the PC valuation function correspond with a MBDC "probably yes" response function. Combined, these results suggest that DC respondents switch from "yes" to "no" nearer to the upper end of the ambivalence region, while PC respondents adopt a more conservative response function.

These results suggest other avenues of future research. Clearly, if the MBDC approach is to be regarded as a viable contingent valuation elicitation format, much additional research will be needed into investigating possible anchoring, centering, ordering, range and other biases. The hypotheses of elicitation-specific preference weighting processes should also be the subject of further research, especially in controlled laboratory experiments in which some "true" reference value is known and in which ambivalence bounds can be measured. The observation that PC and DC values deviate raises the inevitable question of which format provides more valid measures of WTP, and also complicates simulated market comparisons because the reference markets may also be subject to elicitation effects. Given this possibility, it may be necessary for future contingent valuation validity studies to borrow from economic experiments using incentive compatible, demand revealing mechanisms or lotteries in which expected values are known.

References

Arrow, K., 1986. "Review Panel's Assessment", pp. 180-185 in Cummings *et al.*

Arrow, K., R. Solow, E. Leamer, P. Portney, R. Randner, and H. Schuman, 1993. "Report of the NOAA Panel on Contingent Valuation", Appendix I, **Federal Register** 58(10):4602-4603.

Ballestreri, E., G. McClelland, G. Poe, and W. Schulze. "Contingent Values versus Auction Values: An Experimental Comparison of Elicitation Mechanisms", Department of Agricultural, Resource, and Managerial Economics, Cornell University.

Bergstrom, J. C. and J. Stoll, 1989. "Application of Experimental Economics Concepts and Precepts to CVM Field Survey Procedures", *Western Journal of Agricultural Economics*, 14(1):98-109.

Boyle, K. J. and R. C. Bishop, 1988. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques", *American Journal of Agricultural Economics* 70(1):20-28.

Cameron, T. A. and D. D. Huppert, 1989. "OLS versus ML Estimation of Non-market Resource Values with Payment Card Interval Data", *Journal of Environmental Economics and Management*, 17:230-246.

Carson, R. T., R. C. Mitchell, W. M. Hanemann, R. J. Kopp, S. Presser, and P. A. Ruud, 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill", A Report to the Attorney General of the State of Alaska, Vols. 1 and 2, November 10.

Cooper, J. and J. Loomis, 1992. "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models", *Land Economics* 68(2):211-224.

Coursey, D. L. and W. D. Schulze, 1986. "The Application of Laboratory Experimental Economics to the Contingent Valuation of Public Goods", *Public Choice*, 49:47-68.

Cummings, R. C., D. S. Brookshire, and W. D. Schulze (eds.), **Valuing Environmental Goods: An Assessment of the Contingent Valuation Method**, Rowan & Allanheld Publishers, Totowa, NJ, 1986.

Davis, D. D. and C. A. Holt, 1993. **Experimental Economics**, Princeton University Press.

Desvousges, W., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson and K. N. Wilson, 1992. **Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy**, Research Triangle Institute Monograph 92-1, Exxon Corporation.

DuBourg, W.R., M.W. Jones-Lee, and G. Loomes, 1994. "Imprecise Preferences and the WTP-WTA Disparity", *Journal of Risk and Uncertainty* 9:115-133.

Freeman, A. M. III, 1993. **The Measurement of Environmental and Resource Values: Theory and Methods**. Resources for the Future, Washington, DC.

Gregory, R., S. Lichtenstein, and P. Slovic, 1993. "Valuing Environmental Resources: A Constructive Approach", *Journal of Risk and Uncertainty* 7:177-197.

Grether, D. M. and C. R. Plott, 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon", *The American Economic Review* 69(4):623-638.

Hanemann, M., J. Loomis, and B. Kanninen, 1991. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation", *American Journal of Agricultural Economics* 72(4):1255-1263.

Hanemann, W. M., 1995. "Some Issues in Continuous- and Discrete-Response Contingent Valuation Studies," *Northeastern Journal of Agriucltural and Resource Economics* 14(19):5-13.

Hanneman, W. M., 1984. "Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses", *American Journal of Agricultural Economics*, 66(2):332-341.

Hausman, J. A. (ed.), 1993. **Contingent Valuation: A Critical Assessment**. Elsevier Science Publ., Amsterdam.

Hoehn, J.P. and A. Randall, 1987. "A Satisfactory Benefit Cost Indicator from Contingent Valuation", *Journal of Environmental Economics and Management* 14:226-247.

Irwin, J. R., P. Slovic, S. Lichtenstein, and G. H. McClelland, 1993. "Preference Reversals and the Measurement of Environmental Values", *Journal of Risk and Uncertainty* 6:5-18.

Jackson, J., 1965. "Structural Characteristics of Norms," in I. D. Stein and M. Fishbein (eds.), **Current Studies in Social Psychology**, 301-309.

Kagel, J. H. and A. E. Roth (eds.), 1995. **The Handbook of Experimental Economics**, Princeton University Press.

Kanninen, B. J., 1995. "Bias in Discrete Response Contingent Valuation," *Journal of Environmental Economics and Management* 28(1):114-125.

Kanninen, B. J., 1993. "Design of Sequential Experiments for Contingent Valuation Studies", *Journal of Environmental Economics and Management* 25(1):S-1-S-11.

Kealy, M. J. and R. W. Turner, 1993. "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations", *American Journal of Agricultural Economics* 75(2):321-331.

Li, C.-Z. and L. Mattsson, 1995. "Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation", *Journal of Environmental Economics and Management* 28(2):256-269.

Lichtenstein, S. and P. Slovic, 1973. "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas", *Journal of Experimental Psychology* 101(1): 16-20.

Lichtenstein, S. and P. Slovic, 1973. "Reversals of Preference Between Bids and Choices in Gambling Decisions", *Journal of Experimental Psychology* 89(1):46-55.

Magat, W. A., W. Kip Viscusi, and J. Huber, 1988. "Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation", *Journal of Environmental Economics and Management* 15:395-411.

March, J. G., 1978. "Bounded Rationality, Ambiguity, and the Engineering of Choice", *The Bell Journal of Economics* 9(2):587-608.

McFadden, D., 1994. "Contingent Valuation and Social Choice", *American Journal of Agricultural Economics*, 76(4):689-708.

Opaluch, J. J., 1984. "Valuing Natural Resource and Environmental Amenities: Can Economic Valuation Techniques Be Made Defensible: Discussion", *Northeastern Journal of Agricultural and Resource Economics* 13(2):138-141.

Opaluch, J. J. and K. Segerson, 1989. "Rational Roots of 'Irrational' Behavior: New Theories of Economic Decision-Making", *Northeastern Journal of Agricultural and Resource Economics* 18(2):81-95.

Park, T., J. B. Loomis, and M. Creel, 1991. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies," *Land Economics*, 67(1):64-73.

Payne, J. W., J. R. Bettman, and E. J. Johnson, 1992. "Behavioral Decision Research: A Constructive Processing Perspective", *Annual Review of Psychology* 43:87-131.

Poe, G. L., E. K. Severance-Lossin, and M. P. Welsh, 1994. "Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach," *American Journal of Agricultural Economics* 76(4):904-915.

Ready, R. C., J. C. Whitehead, and G. C. Bloomquist, forthcoming. "Contingent Valuation when Respondents are Ambivalent", *Journal of Environmental Economics and Management.*

Schkade, D. A. and J. W. Payne, 1993. "Where Do the Numbers Come From? How People Respond to Contingent Valuation Questions", in J.A. Hausman. pp. 271-293.

Shelby, B. 1981. "Encounter Norms in Backcountry Settings: Studies of Three Rivers," *Journal of Leisure Research*, 13(2):129-138.

Shelby, B., T. C. Brown, and R. Baumgartner, 1992. "Effects of Streamflows on River Trips on the Colorado River in the Grand Canyon, Arizona," *Rivers*, 3(3):191-201.

Slovic, P. and S. Lichtenstein, 1983. "Preference Reversals: A Broader Perspective", *American Economic Review* 73(4):596-605.

Smith, V. L., 1982. "Microeconomic Systems as an Experimental Science", *American Economics Review* 72(5): 923-955.

Smith, V. L., 1987. "Experimental Methods in Economics", in J. Eatwell, M. Milgate, and P. Newman (eds.), **The New Palgrave: A Dictionary of Economics**, MacMillan Press, London, Vol. 2, pp. 241-249.

Svento, R., 1993. "Some Notes on Trichotomous Choice Discrete Valuation", *Environmental and Resource Economics*, 3:533-543.

Tversky, A., S. Sattath, and P. Slovic, 1988. "Contingent Weighting in Judgment and Choice", *Psychological Review* 95(3):371-384.

Tversky, A., P. Slovic, and D. Kahneman, 1990. "The Causes of Preference Reversal", *The American Economic Review* 80(1):204-217.

Welsh, M. P. and R. C. Bishop, 1993. "Multiple Bounded Discrete Choice Models", in John C. Bergstrom (ed.), **Benefits & Costs Transfer in Natural Resource Planning,** Western Regional Research Publication, W-133, Sixth Interim Report, Department of Agricultural and Applied Economics, University of Georgia, September, pp. 331-352.

Figure 1. Uncertain Preferences, Ambivalence, and WTP

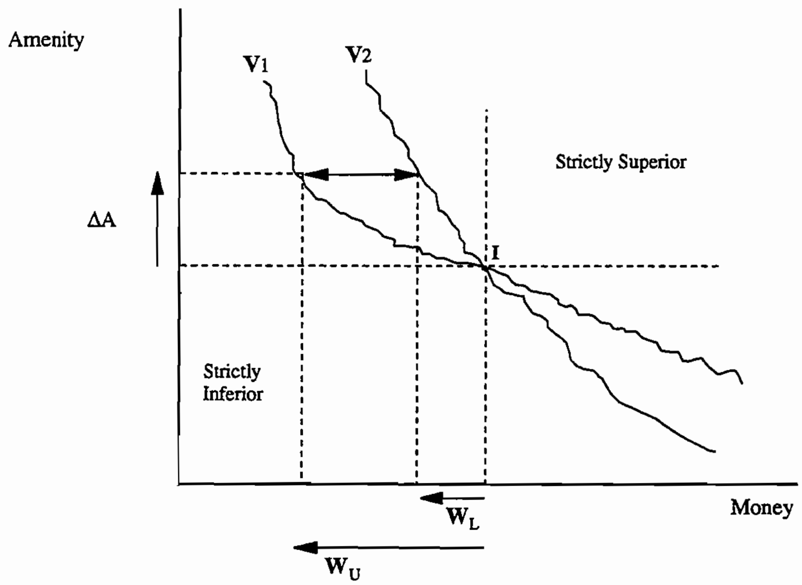


Figure 2. Multiple Bounded/Return Potential Format

Suppose that citizens and residents, including foreign students on visas, would have to make up for part of the revenues lost as a result of this proposal. How would you vote on this proposal? As you think about your answer, please remember that if this proposal passes, you would have less money for personal expenses or to spend on other environmental issues.

3. Would you vote for this proposal if passage of the proposal would cost you these amounts **every year** for the foreseeable future? *(CIRCLE ONE LETTER FOR EACH DOLLAR AMOUNT TO SHOW HOW YOU WOULD VOTE)*

Cost to you per year?	Definitely No	Probably No	Not Sure	Probably Yes	Definitely Yes
10¢	A	B	C	D	E
50¢	A	B	C	D	E
$1	A	B	C	D	E
$5	A	B	C	D	E
$10	A	B	C	D	E
$20	A	B	C	D	E
$30	A	B	C	D	E
$40	A	B	C	D	E
$50	A	B	C	D	E
$75	A	B	C	D	E
$100	A	B	C	D	E
$150	A	B	C	D	E
$200	A	B	C	D	E

Figure 3. Base Scenario, All Versions

A PROPOSAL

Under this proposal, daily fluctuations in the river level would be eliminated. Seasonal releases would also be changed so that releases would be highest during the spring, just like before the dam was built. However, the highest spring releases would still be lower than the average springtime flow prior to the dam's construction. If this proposal is selected, it will result in the following environmental conditions along the Colorado River in the Grand Canyon:

- In the long-term, the number and size of beaches would remain at present levels.
- The risk of erosion to Native American traditional use areas, sacred sites, and archeological sites would decrease substantially.
- The area available for vegetation would increase by about 10% so that the area available for birds and other forms of wildlife would increase by about 10%.
- There would be a major improvement in conditions for native fish. Populations of most native fish, including one of the species in danger of extinction, would increase.
- There would be a major improvement in conditions for trout. The size and number of trout would increase. Maintenance of the trout population would no longer require annual stocking.

2. Think about a situation in which you have an opportunity to vote on this proposal. If passage of this proposal would not cost you anything, would you support this proposal? *(CIRCLE ONE NUMBER)*

   1 **No**---------->*SKIP TO QUESTION 4*
   2 **Yes**
   3 **I would choose not to vote on this proposal**---------->*SKIP TO QUESTION 4*

Figure 4. Payment Card Format

> Suppose that citizens and residents, including foreign students on visas, would have to make up for part of the revenues lost as a result of this proposal. How would you vote on this proposal? As you think about your answer, please remember that if this proposal passes, you would have less money for personal expenses or to spend on other environmental issues.

3. If the passage of the proposal would cost you these amounts **every year** for the foreseeable future, what is the highest amount that you would pay and still vote for the program? (*CIRCLE THE HIGHEST AMOUNT THAT YOU WOULD STILL VOTE FOR THE PROGRAM)*

10¢	50¢	$1	$5	$10	$20
$30	$40	$50	$75	$100	$150
$200	MORE THAN $200				

Figure 5. Dichotomous Choice Format

> Suppose that citizens and residents, including foreign students on visas, would have to make up for part of the revenues lost as a result of this proposal. How would you vote on this proposal? As you think about your answer, please remember that if this proposal passes, you would have less money for personal expenses or to spend on other environmental issues.

3. Would you vote for this proposal if passage of the proposal would cost you $__________ **every year** for the foreseeable future? (*CIRCLE ONE NUMBER)*

    1 **No**

    2 **Yes**

Figure 6. Estimated Logit Distributions

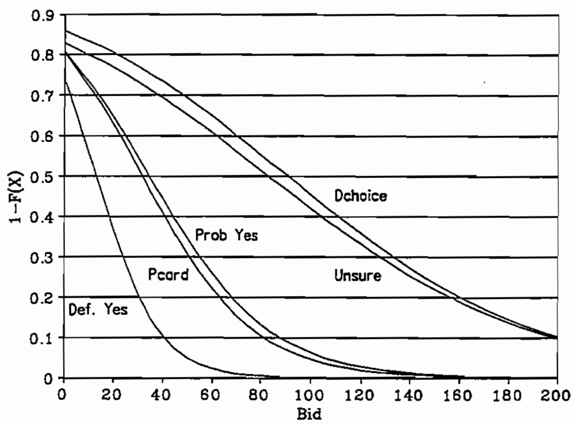


Table 1. Actual Response Distributions

	Multiple Bounded / Return Potential %[1]					PCard (%)	DC (% yes)
	Def. No	Prob. No	Unsure	Prob. Yes	Def. Yes		
10¢	0.0	0.5	0.5	6.8	92.2	2.7	
50¢	0.5	0.5	1.6	9.4	88.0	3.2	
$1	0.5	0.5	4.7	13.6	80.6	9.6	92.0
$5	3.6	3.6	11.5	20.3	60.9	12.8	91.3
$10	6.8	5.7	15.6	29.2	42.7	20.2	95.4
$20	12.5	10.9	18.2	28.1	29.7	10.1	80.0
$30	18.2	13.5	20.8	29.7	17.7	4.8	68.2
$40	25.0	12.0	24.5	27.1	11.5	7.4	
$50	29.7	16.1	27.1	19.8	7.3	14.9	40.9
$75	35.9	18.8	28.6	13.0	3.6	1.6	
$100	46.4	20.8	23.4	6.8	2.1	8.5	56.5
$150	53.6	22.4	18.2	4.2	1.6	1.1	14.3
$200	60.7	17.8	16.8	3.7	1.0	2.7	19.1

[1] For multiple Bounded / Return Potential Format % sum to 100 within rows.

Table 2. Estimated Logit Models [1,2]

	α	β	n
Def. Yes, MBDC	1.048 (0.136)***	-0.080 (0.005)***	185
Prob. Yes, MBDC	1.426 (0.146)***	-0.041 (0.003)***	185
Not Sure, MBDC	1.572 (0.154)***	-0.019 (0.001)***	185
Payment Card	1.404 (0.143)***	-0.044 (0.003)***	188
Dichotomous Choice	1.806 (0.249)***	-0.020 (0.003)***	204

[1] *** denotes 1% significance level.

[2] Using likelihood ratio tests, payment card and probably yes models were not significantly different (LR = 0.93), and dichotomous choice and not sure models were not significantly different (LR = 0.95) at the 5 percent significance level. The null hypothesis of equality was rejected for all other models. ($\chi^2_{0.05,2} = 5.99$)

Table 3. Krinsky and Robb Simulation Values for Median ($-\alpha/\beta$) and Mean of Non-Negative Distribution

	Median at Parameters [95 Percent CI]	NN Mean at Parameters [95 Percent CI]
MultA	12.96 [9.92,16.34]	16.70 [14.33,19.51]
MultB	34.38 [28.87,39.80]	39.56 [34.62,44.40]
MultC	82.96 [70.05,96.91]	92.96 [81.10,106.15]
Payment Card	31.69 [26.02,37.18]	36.64 [32.05,41.64]
Dichotomous Choice	90.76 [72.38,112.38]	98.40 [81.56,123.86]

Significance (α) of difference: DC - MultC — Median α=0.53; NN Mean α=0.66

MultB - PCard — Median α=0.55; NN Mean α=0.45

All other combinations did not overlap, and the distribution of the convolution did not include zero.

Sensitivity of Multiple-Bound Referendum CV Estimates to the Specification of the Underlying Utility

by

Anna Alberini
Fellow, Resources for the Future
Washington, D.C.

and

Joseph Cooper
Economist, Economics Research Service
U.S. Department of Agriculture
Washington, D.C.

Sensitivity of Multiple-Bound Referendum Contingent Valuation Estimates to the Specification of the Underlying Utility

by

Anna Alberini and Joseph Cooper

I. Introduction

In the absence of markets for environmental quality and amenities, the travel cost method, hedonic equation approaches and contingent valuation surveys have been used extensively to estimate the changes in welfare associated with environmental policies, public goods and amenities. In practice, such welfare changes will depend on the assumptions about individual preferences and on the econometric model fit to the data.

Using artificially generated travel cost data, for instance, Kling (1988, 1991) finds that the point estimates of welfare change vary with the specific econometric model one chooses to fit. Drawing on data on housing prices in the Baltimore area, Cropper et al. (1993) simulate transactions between individuals in the local housing market until the market clears, fit hedonic and multinomial logit models to the artificial transaction data, and find that the performance of the model changes with changing the "true" model, omitting regressors or replacing them with proxies.

Unfortunately, the findings about the impact of specification choices from the travel cost literature and hedonic equation literature are not readily extended to contingent valuation, because of the different nature of the data and models involved. Most recent contingent valuation surveys (listed in Carson et al., 1995) elicit information on willingness to pay (WTP) in the form of a "yes"/"no" response to a dichotomous choice payment question and, accordingly, fit binary data models. Count data models and truncated data models, for instance, are commonly used with travel cost data to accommodate the format of the dependent variable (number of trips to a site taken by an individual), and the tendency to observe "corner solutions" (*i.e.*, the fact that an individual may never visit certain sites at all).

The estimates of welfare change obtained from CV survey data are also likely to vary with the assumptions about preferences, *i.e.* with the shape of the individual's utility function. Econometric specification decisions and assumptions on individual preferences are usually tightly intertwined when dealing with CV survey data: Hanemann (1984) shows that the structure of the utility function dictates what regressors should be included in the econometric models of dichotomous choice CV data and what their functional forms should be.

In this paper attention is restricted to dichotomous choice contingent valuation. We are specifically concerned with how econometric specification and assumptions on individual preferences reflect on the performance of the model and on WTP estimates. We run Monte Carlo simulations in which we generate artificial dichotomous choice "contingent valuation" data from a variety of "true" models and fit econometric models commonly used in CV practice. We find that coefficients and welfare estimates vary dramatically with the fitted model, to the point that minor specification changes can result in large biases of the estimates. We also look at regression diagnostics for guidance in selecting between models, but find that many of them are not particularly informative.

This paper is organized as follows. We briefly review the economics and the econometrics of dichotomous choice contingent valuation data and explain the purpose of our investigation in Section II. The experimental design is described in Section III. The experimental results are discussed in Section IV. Section V concludes.

II. Dichotomous Choice Contingent Valuation

In a dichotomous choice CV survey a respondent is asked to state whether he would vote in favor or against a proposed environmental policy that -- if approved by majority -- would be implemented at the cost of $X to the respondent's household. The response to this question can be interpreted in two possible ways.

Following Hanemann (1984), the random-utility theoretic interpretation postulates that the "yes"/"no" response is motivated by the utility associated with the provision of the plan (the respondent's disposable income being reduced by $X) compared with the utility of the status quo. Utility is, in turn, decomposed into a deterministic and a stochastic part, the latter reflecting a component of taste that is not observable by the researcher. Hanemann derives logit and probit equations and expressions for the welfare change, depending on the distribution of the error term and on the structure of the deterministic portion of utility. The econometric model is a logit equation if the difference between the random components of utility (with and without the policy) is distributed as a logistic, and a probit if the difference in random components is normally distributed. If the deterministic portion of utility is linear in income the probit/logit equation includes at the right-hand side the intercept and the cost amount stated to the respondent in the survey. The logit/probit equation changes to include the intercept and a regressor consisting of the cost amount divided by the respondent's income if the deterministic component of utility is log linear in income.[1]

Another popular approach, illustrated by Cameron and James (1987), is based on the distribution of the underlying willingness to pay directly. This approach recommends running probit or logit regressions of the responses on the intercept, individual characteristics and the bid value, or a suitable transformation of it. In at least one important case, however, the two approaches reduce to the same econometric model.[2]

Because WTP estimates from contingent valuation studies are often used for policy purposes and in litigation over natural resource damage (see Kopp and Smith, 1993), it is important to determine how heavily they are affected by the interaction of econometric assumptions (*e.g.*, the distribution of the error term; the mapping between the underlying quantities and the observables) and assumptions on preferences (*e.g.*, the shape of the utility function, if explicit; the variables that influence willingness to pay). How much should the

---

[1] As an additional example, if the utility function is quadratic in income the probit/logit equation can be shown to include the intercept, the stated cost amount, the *square* of the stated cost amount and an *interaction* term between the cost amount and the respondent's income at the right-hand side.

[2] McConnell (1990) shows that the approach proposed by Cameron and James (1987) is also utility-driven and that it starts from the expenditure function.

investigator be guided by *a priori* choices of the utility function and by empirical considerations such as fit and significance of the estimated coefficients?

To make the answers to these questions even more complicated, recent contingent valuations surveys have included follow-up payment questions and the standard probit/logit models have been extended to form interval-data (or "double-bounded") models.[3] Hanemann, Loomis and Kanninen (1991) show that the estimates obtained using the "double-bounded" protocol are statistically more efficient than their single-bounded counterparts. However, if variables are omitted from the econometric model or entered in the wrong functional form, it is possible that these misspecifications affect the WTP estimates more strongly in a double-bounded model than in a single-bounded model, due to the heavier reliance on the maintained assumptions effectively enacted by double-bounded models. Double-bounded models assume that the responses to both payment questions are generated by the same, unobserved willingness to pay amount.[4]

In addition to the fundamental goal of obtaining estimates of welfare change, usually in the form of mean or median willingness to pay, in practice many contingent valuation studies are conducted for the purpose of obtaining estimates for a particular coefficient (or a function of it). This is often the case when portions of the sample are administered variants of the CV that alter an attribute of the policy (as with a "scope test;" see Arrow et al, 1993), or when estimates of the elasticity of WTP with respect to an economic quantity, such as income, are desired.

Researchers may also care for high-quality predictions of the curve that gives --for any given cost -- the fraction of the population willing to vote in favor of the policy. In this case special attention would be devoted to the *predicted* "yes" or "no" responses. In addition, predictions are often compared with actual responses to assess the fit of the model and as a

---

[3] If a follow-up is used, a respondent who votes in favor of the plan at cost \$X is asked to state what his vote would be if the cost were \$(X+Δ) (with Δ>0). A lower cost \$(A-Δ) is offered to a respondent who favors the status quo in the initial choice.

[4] Alberini, Kanninen and Carson (1994) propose and estimate models in which the latent willingness to pay amount changes after the first payment question, reflecting respondent confusion and strategic consideration occurring with the follow-up question.

criterion of model selection in the presence of competing models. This has prompted us to check whether empirical models that offer reasonably good predictions for the responses also give valid WTP estimates.

III. Experimental Design

To find out how the specification and preference decisions affects the results of the study, we generate artificial dichotomous choice CV data, fit single- and double-bounded models (see Hanemann and Kanninen, forthcoming, for a survey of single- and double-bounded models and issues) and obtain coefficients, WTP estimates and predictions for the individual responses. Because we fit the correct models as well as other (incorrect) specifications, our investigation allows us to find out whether fitting "wrong" models necessarily provides biased benefit estimates, whether poor specification choices can be discovered using diagnostic checks, and whether certain specifications are more "robust" to misspecifications than others.

We run a complete Monte Carlo simulation exercise in which we draw hundreds of artificial samples of WTP amounts to obtain distributions of estimated coefficients, welfare measures and predictions. For the purposes of this paper, attention is concentrated on (i) one welfare measure, median WTP, which in many cases coincides with mean WTP (see below); (ii) coefficients from the model, plus one statistic that has economic interpretation, the elasticity of willingness to pay with respect to income; and (iii) two sets of predictions based on the fitted model.

Table 1 summarizes our experimental design. We generate artificial CV responses from six models. Two of these models (models [1] and [2]) follow from the random utility framework proposed by Hanemann (1984) and assume normally distributed errors. Model [1] assumes that utility depends on log income,[5] whereas model [2] assumes that utility is linear in income. Model [2] coincides with assuming a normally distributed willingness to pay. Two

[5] Because income is retained in the econometric specification corresponding to this utlity function, the associated probit equation is referred to as the "income effects" model in the remainder of this paper.

more models (models [3] and [4]) imply a normally distributed WTP, with mean WTP involving income and log income, respectively. For these models, mean and median WTP are identical. Finally, two models (models [5] and [6]) both imply a log normal WTP, but differ in the sense that (log) income is included as a covariate in [6] but not in [5].

As shown in the third column of Table 1, we fit probit equations corresponding to three possible models for latent WTP: (i) a linear-in-income utility function (or, a normally distributed willingness to pay); (ii) the "income effects" model; and (iii) a log normal WTP.[6] We include income at the right-hand side only when income is indeed used in generating the data. Income is entered in a linear fashion when WTP is assumed to be normally distributed, and after taking its log transformation if the fitted model assumes WTP to have a log normal distribution.

We consider three possible ways of generating the responses to the follow-up questions. We first assume that the responses to the initial and follow-up questions are driven by one WTP amount (*i.e.*, the WTP amount does not change between questions). We then move to the situation in which the response to the follow-up question is driven by a WTP amount that is linearly dependent with, but not identical to, the WTP amount underlying the initial response, their correlation coefficient being 0.5. The last set of simulations assumes that the two responses are driven by two independent WTP amounts, but that the distributions of the two WTP variables are identical.

[6] The linear-in-income utility function requires fitting a probit model in which the bid is entered linearly at the right-hand side. In the "income effects" model a variable consisting of bid divided by income is included at the right-hand side. When a log normal willingness to pay is assumed, the log transformation of the bid is entered at the right-hand side of the probit model of the responses.

Table 1. Experimental Design: True Data Generation and Fitted Models.

True Model Used to Generate Data	True welfare measure	Correct Econometric Model Probit Regression on:	Fitted Econometric Models. Probit Regression on:
[1] Random Utility Model with U=α+β*log(income)+error	$151.04	constant, bid/income [True parameter values: const=0.4300, bid/income=-64.6200]	1. constant, bid, income 2. constant, bid/income 3. constant, log bid, log income
[2] Random utility model with U=α+β*income+error (also interpreted as a WTP model with WTP=A+error)	Mean/median WTP: $76.25	constant, bid [True parameter values: const=0.5750, bid=-0.00754]	1. constant, bid 2. constant, bid/income 3. constant, log bid
[3] WTP=A+B*income+error	Mean/median WTP: $132.39	constant, bid, income [True parameter values: const=0.0950, bid=-0.0076, income=0.0004]	1. constant, bid, income 2. constant, bid/income 3. constant, log bid, log income
[4] WTP=A+B*log income+ error	Mean/median WTP: $76.35	constant, bid, log income [True parameter values: const=-0.50468, bid=-0.00758, log income=0.11162]	1. constant, bid, income 2. constant, bid/income 3. constant, log bid, log income
[5] log WTP=A+error	Median WTP: $53.99	constant, log bid [True parameter values: const=0.8600, log bid=-0.21560]	1. constant, bid 2. constant, bid/income 3. constant, log bid
[6] log WTP=A+B*log income+ error	Median WTP: $55.65	constant, log bid, log income [True Parameter values: const=0.0800, bid=-0.21794, log income=0.0820]	1. constant, bid, income 2. constant, bid/income 3. constant, log bid, log income

We use sample sizes of 250, 500 and 1000, and performed 500 replications for each set of simulations. Our artificial respondents are evenly and randomly divided among the four bid sets reported in Table 2. These bid sets follow closely those used in the CV survey conducted to value the natural resources of Kakadu Conservation Zone/Park in Australia (Carson, Imber and Wilks, forthcoming). The income values attributed to our artificial respondents are randomly drawn the distribution of income in the sample from the Kakadu CV survey. The values of the parameters used in generating the data were obtained after fitting similar models to the subsample administered the "minor impact" scenario in the Kakadu CV survey.

Table 2. Bid Values Used in the Simulations

Bid Set	Initial Bid	Lower Follow-up Bid	Higher Follow-up Bid
1	5	2	20
2	20	5	50
3	50	20	100
4	100	50	250

IV. Results

Welfare Statistics

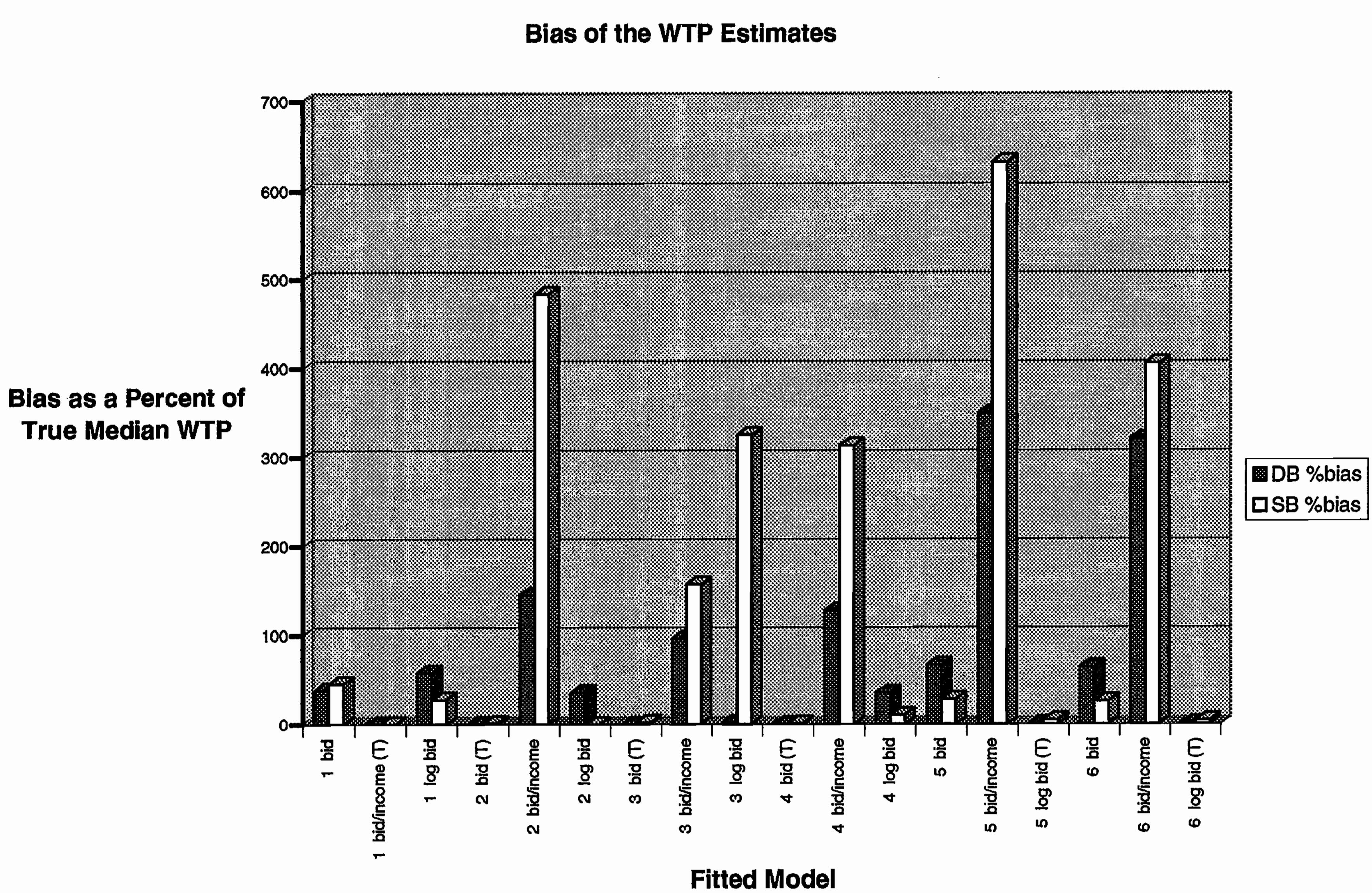
Our comments are limited to the results of the sets of simulations with sample size 1000. We begin our analysis of the results with the case in which one WTP amount drives the responses to both the initial and the follow-up questions. The relevant figures and tables are labeled as having coefficient of correlation equal to one.

The biases of median WTP observed in the simulations (expressed as percent of true median WTP; signs are neglected) are depicted in Figure 1 for all fitted models. The true model is indicated by its number from Table 1, while the symbol (T) marks the correct econometric model.

The figure shows clearly that the biases are always relatively small -- in fact, negligible -- when the correct econometric model is fit. The double-bounded estimates, in particular, tend to have smaller biases than the single-bounded estimates obtained neglecting the responses to the follow-up payment questions: the biases of the double-bounded estimates are no greater than 1% of the true median WTP, whereas the biases of the single-bounded estimates of median WTP get to be as large as 5 percent of the true median.

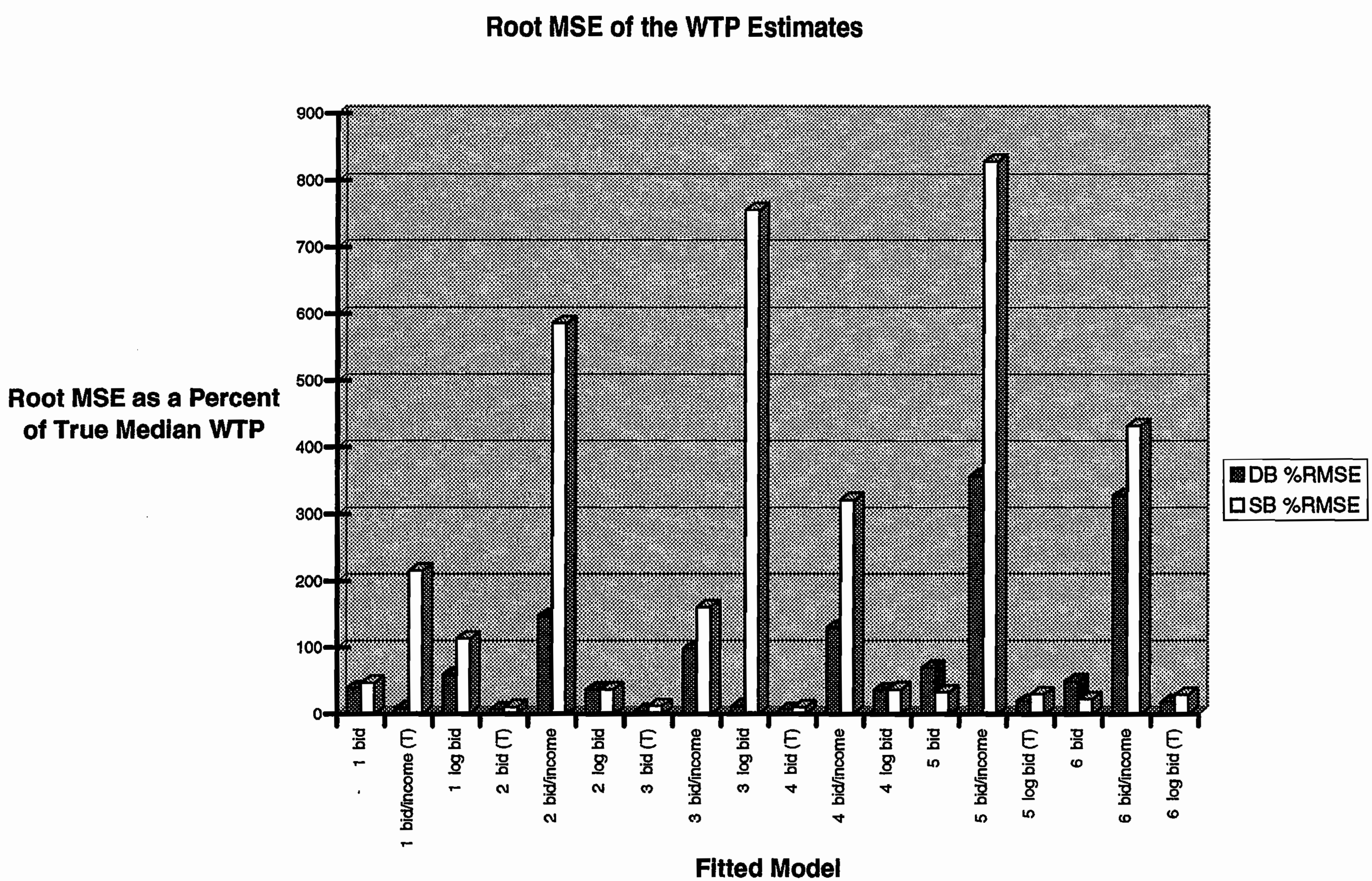
Fitting "wrong" models, however, typically results in much larger biases. The biases appear to be especially large when the "income effects" utility model is incorrectly chosen. In that case the double-bounded estimates of median WTP range between one and three times the true median WTP, depending on the true and fitted models, whereas the single-bounded estimates range between 1.5 and over six times the true median WTP.

Figure 1



Bias of the WTP Estimates
Bias as a Percent of True Median WTP
0
100
200
300
400
500
600
700
1 bid
1 bid/income (T)
1 log bid
2 bid (T)
2 bid/income
2 log bid
3 bid (T)
3 bid/income
3 log bid
4 bid (T)
4 bid/income
4 log bid
5 bid
5 bid/income
5 log bid (T)
6 bid
6 bid/income
6 log bid (T)
Fitted Model
DB %bias
SB %bias

Figure 2


Root MSE of the WTP Estimates
Root MSE as a Percent of True Median WTP
900
800
700
600
500
400
300
200
100
0
1 bid
1 bid/income (T)
1 log bid
2 bid (T)
2 bid/income
2 log bid
3 bid (T)
3 bid/income
3 log bid
4 bid (T)
4 bid/income
4 log bid
5 bid
5 bid/income
5 log bid (T)
6 bid
6 bid/income
6 log bid (T)
Fitted Model
DB %RMSE
SB %RMSE

Another misspecification considered in our simulations is to assume a normal distribution for WTP when the true distribution is a log normal, and viceversa (models [3] to [6]). Figure 1 shows that biases resulting from this type of misspecification can also be quite serious, although generally to a lesser degree than those incurred when the "income effects" model is fit. The most serious bias is that observed when the true model is [3] (normal WTP that is linear in income), but the model that assumes a log normal distribution is fitted, and the analysis of the data is limited to the responses to the initial payment questions. In that situation we find that the bias is about 300 percent of the true median WTP. Fortunately, the bias drops substantially when the responses to the follow-up payment questions are used. Even in the other cases entailing wrong distributional assumptions, however, the biases are generally not negligible: they range from one-third to two-thirds of true median WTP.

On comparing results for all true models, we feel it is safe to conclude that there is no evidence that "wrong" fitted models result in larger biases with double-bounded estimation procedures. In some cases the double-bounded estimates are biased less heavily than the single-bounded, in other cases this relationship is reversed. The single-bounded estimates are always biased more heavily than the single-bounded estimates when the "income effects" model is fitted. In the examples with truly large biases the single-bounded procedure usually does worse than the double-bounded estimation procedure. The single-bounded estimates are often less biased than the double-bounded estimates in the examples with moderate biases.

Figure 2 displays information about the mean square error (MSE) of the estimates. The MSE (the sum of bias square and variance of the estimates) is a useful indicator of the tradeoff between unbiasedness and precision, and is here calculated as:

$$(1) \qquad MSE = \frac{1}{500}\left[\sum_{j=1}^{500}(\hat{\theta}_j - \theta)^2 + \sum_{j=1}^{500}(\hat{\theta}_j - \bar{\theta})^2\right]$$

where $\hat{\theta}_j$ is the estimate of median WTP obtained from the j-th replication, θ is the true median WTP, and $\bar{\theta}$ is the average of the estimated median WTP.

Rather than giving the "raw" MSE, we display the square root MSE, expressed as a percent of true median WTP, in Figure 2. The figure shows clearly that when the true model is fitted the MSE tends to be small. It is smaller for double-bounded estimates because of their smaller biases and higher precision than the single-bounded estimates. Under correct specification, the variance of the estimates accounts for most of the MSE.

When "wrong" models are fit, the MSE can become very large. Figure 2 shows that in these cases root MSE can be several times the true median WTP. It is now the bias to account for most of the MSE, especially for the double-bound estimates.

Since incorrect distributional assumptions and functional forms for bid and covariates are shown to result in potentially serious biases, one is left wondering whether, holding the fitted econometric model unchanged, the changes in the WTP estimates occurring when moving from single-bounded to double-bounded models can help detect misspecifications. Although those changes can indeed be very large (see Figure 1), we find that only in 2 cases out of 18 is the difference between the average single-bounded median WTP and the average double-bounded median WTP significant. This finding is explained primarily by the fact that the biases of single-bounded and double-bounded estimates are often of the same sign and by the extremely large variances of the single-bounded estimates, which do not allow to discriminate between single- and double-bounded estimates. This finding also suggests that, for misspecifications of the types considered in this paper, in many cases a Hausman test will fail to reject the null of a significant difference between single-bounded and double-bounded WTP estimates, thus leading researchers to place confidence in poor assumptions and estimates.[7]

Non-parametric Techniques

Researchers might resort to non-parametric techniques for estimating willingness to pay in order to avoid grossly biases. One approach might be to calculate a non-parametric estimates

[7] The Hausman test might be more suitable to detect differences between single- and double-bounded estimates caused by other factors, such as undue sensitivity of the WTP responses to the bid amounts.

of the survival function of willingness to pay (see Lee, 1992) and to calculate the area under this step function. One shortcoming of this approach is that the estimates of mean WTP tend to be sensitive to the highest bid value used.

The recent *Montrose* CV study (NRDA, Inc., 1994) relies on a conservative, non-parametric procedure to produce a lower-bound estimate for mean WTP. Following that approach, we form intervals around the respondents' WTP amounts using the responses to the initial and follow-up payment questions, and then impute the lower bound of these intervals as the respondent's WTP amount. We finally calculate the sample mean of these imputed WTP values.

Table 3 compares the lower-bound sample means, averaged over the replications, with the estimates of median WTP based on the fitted parametric models, also averaged over the replications. While it is not surprising that the lower-bound mean is always much smaller than the true median WTP and any parametric WTP estimate, we point out that the *distance* between the lower-bound estimate and the parametric WTP estimates does not convey information that can help uncover which model is correct.[8]

[8] The distance between lower-bound estimate and the parametric estimate is not necessarily smallest when the parametric model is the correct model.

Table 3. Parametric and Non-parametric Estimates of Welfare Change.
(Simulations with sample size = 1000 and correlation coefficient = 1.)

True Model	Fitted Model	Median WTP Averaged over the Replications (in $)	Lower-bound Mean WTP Averaged over the Replications (in $)
[1]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	132.38 259.20 129.67	54.8483
[2]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	93.07 150.39 63.09	47.1093
[3]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	76.01 186.66 49.09	40.3005
[4]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	90.24 242.64 54.52	47.9673
[5]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	76.17 173.70 49.33	40.3015
[6]	1. const, bid, income 2. const, bid/income 3. const, log bid, log income	91.74 234.13 56.10	48.2585

Coefficients

Descriptive statistics for the coefficients of bid and income (or log income), when the latter is included in the model, are displayed in Table 4. The mean of the distribution of the bid coefficient is significantly different from zero in all models of this paper. (This statement is true of the coefficient of bid divided by income for the "income effects" utility model). This finding is based on the value of the asymptotic "t-statistic" obtained dividing the average value of the coefficients of the bid divided by the standard deviation of these coefficients. Interestingly, the t-statistic computed in this fashion attains the largest value -- in absolute magnitude -- when the fitted model *is* the correct model. Accordingly, a possible strategy for model selection might indeed be to look for the model with the most strongly significant bid coefficient. We warn the reader, however, that sampling variability may well result in the correct econometric model having a lower t-statistic for the bid coefficient than competing models. Furthermore, the set of simulations we carried out is not comprehensive enough to give a conclusive answer about the t-statistic for the coefficient of the bid.

The coefficient of the income variable is also generally significant, with the only exception of true model [6]. We also look at a function of this coefficient, *i.e.* the income elasticity of willingness to pay (Table 5). Once again, moving from one fitted model to another results in dramatic changes (even 100%) in the income elasticity of willingness to pay. This suggests that researchers should be prepared to obtain very different elasticities of willingness to pay with respect to income as they change the model specification.

Table 4. Descriptive Statistics for Model Parameters.
(Simulations with sample size = 1000 and correlation coefficient = 1.)

True Model	Fitted Model	Average Value of the Coefficient of Bid over the Replications (standard deviation in parenthesis)[a]	Average Value of the Coefficient of Income over the Replications (standard deviation in parenthesis)
[1]	1. const, bid, income	-0.0042 [0.00039]	1.1800e-05 [2.6041e-06]
	2. const, bid/income	-64.9164 [4.23468]	N/A
	3. const, log bid, log income	-0.18871 [0.01381]	0.2622 [0.0418]
[2]	1. const, bid	-0.00755 [0.00041]	
	2. const, bid/income	-48.3640 [5.22780]	
	3. const, log bid	-0.30633 [0.01747]	
[3]	1. const, bid, income	-0.0076 [0.00043]	4.0258e-05 [3.3974e-06]
	2. const, bid/income	-64.2117 [13.6242]	N/A
	3. const, log bid, log income	-0.3402 [0.02283]	0.5489 [0.0464]
[4]	1. const, bid, income	-0.00755 [0.00410]	
	2. const, bid/income	-54.4663 [5.96060]	
	3. const, log bid, log income	-0.30674 [0.01738]	
[5]	1. const, bid	-0.00429 [0.00029]	
	2. const, bid/income	-27.9222 [3.22596]	
	3. const, log bid	-0.21570 [0.01400]	
[6]	1. const, bid, income	-0.00430 [0.00030]	3.5840e-06 [2.3601e-06]
	2. const, bid/income	-30.2163 [3.52000]	N/A
	3. const, log bid, log income	-0.2182 [0.01390]	0.0841 [0.0429]

[a] Table refers to the coefficient of bid/income for the "income effects" probit model (fitted model 2.).

Table 5. Income Elasticities of Willingness to Pay.
(Simulations with sample size = 1000 and correlation coefficient = 1.)

True Model	True Income elasticity of WTP (evaluated at the sample mean of the regressors)	Fitted Model	Income elasticity of WTP averaged over the replications	Standard Deviation of income elasticity of WTP
[1]	N/A	1. const, bid, income 2. const, log bid, log income	0.6900 1.4000	0.1500 0.2400
[3]	0.99	1. const, bid, income 2. const, log bid, log income	0.9100 1.6193	0.0700 0.1609
[6]	0.37625	1. const, bid, income 2. const, log bid, log income	0.20775 0.37037	0.13801 0.20124

Predictive Performance

Turning to measures of fit and predictive performance, we obtain predictions in two ways. For models that include income (or log income) at the right-hand side, we obtain predictions for WTP (log WTP) as $WTP_i = \hat{a} + \hat{b} * y_i$ ($\log WTP_i = \hat{a} + \hat{b} * \log y_i$), where $\hat{a}$ and $\hat{b}$ are the estimates from the double-bounded model,[9] and predict a "yes" response if predicted WTP for the i-th respondent is greater than the bid (log bid) assigned to that respondent. We repeat the procedure at the follow-up bid to generate the predicted response to the follow-up question.[10] We then compare the pair of responses predicted for that individual with the actual pair of responses and count the proportion of correctly predicted YY, NN, YN and NY pairs.

Table 6 reports the average proportion of predicted responses. There is very little difference between the correct model and a model based on wrong distributional assumptions in terms of the proportion of correctly predicted responses, suggesting that this notion of predictive performance offers little help in detecting misspecifications. We also point out that even the

[9] These estimates are obtained as $\hat{a}$=-(intercept of probit model/coefficient of bid) and $\hat{b}$=-probit coefficient of income/coefficient of bid). The coefficients of log bid and log income are used in the probit model that assumes a log normal willingness to pay.

[10] To summarize, we predict a YY if predicted WTP_i (log WTP_i) is greater than both the initial bid and the follow-up (their logs), a YN if predicted WTP_i (log WTP_i) is greater than the initial bid but lower than the follow-up (their logs), a NY if predicted WTP_i (log WTP_i) is less than the initial bid but greater than the follow-up amount (their logs), and finally a NN if predicted WTP_i (log WTP_i) is less than both the initial and follow-up amounts.

correct model does not necessarily predict well certain pairs of responses. The proportion of correctly predicted YY and NN pairs is much lower than the proportion of correctly predicted YN and NY, probably as a result of our experimental design.

The second measure of predicted performance is a Pearson chi square test, which we use with the models without covariates. The Pearson chi square we compute compares the predicted "yes" and "no" responses at a stated value of the bid (we choose the initial bid values) with the actual counts. The difference between predicted responses (the probability of "yes" or "no" at a given bid level obtained from the distribution of WTP, times the number of people assigned that initial bid level) and actual is squared and standardized by the number of predicted "yes" or "no" responses:

$$(2) \qquad \sum_{k=1}^{K}\left[\frac{(n_k\cdot \Pr(yes)-m_k)^2}{n_k\cdot \Pr(yes)}+\frac{(n_k\cdot(1-\Pr(yes))-(n_k-m_k))^2}{n_k\cdot(1-\Pr(yes))}\right]$$

where n_k is the number of respondents assigned the k-th initial bid level and m_k is the count of positive responses at the k-th initial bid level. We use the double-bounded estimates of the parameters of the cdf of WTP to calculate the probabilities in expression (2). The test is distributed as a chi square with two degrees of freedom if the specification is adequate (giving no statistical difference between predicted and actual counts of "yes" and "no" responses).

The results of the Pearson chi square test are reported in Table 7. At least for the cases we considered, it appears that the distributional and functional form misspecifications are picked up by the test, although we are somewhat concerned about the size of the test being higher than the nominal size. We remind the reader, however, that the Pearson chi square test is not very powerful for small sample sizes (Alberini, 1995).

Table 6. Predictive Performance of the Models with Covariates.
(Simulations with sample size = 1000 and correlation coefficient = 1.)

True Model	Fitted Model	Percent Correctly Predicted YY	Percent Correctly Predicted NN	Percent Correctly Predicted YN	Percent Correctly Predicted NY
[1]	1. const, bid, income	61.35	51.64	91.01	95.03
	2. const, log bid, log income	64.12	59.93	90.99	95.05
[3]	1. const, bid, income	67.28	74.46	85.85	94.32
	2. const, log bid, log income	65.88	74.31	85.85	94.32
[6]	1. const, bid, income	53.07	62.51	91.83	92.09
	2. const, log bid, log income	55.74	64.04	91.83	93.10

Table 7. Predictive Performance for Models without Covariates.
(Simulations with sample size = 1000 and correlation coefficient = 1.)

True Model	Fitted Model	Pearson Chi Square Test Averaged over the Replications	Relative Frequency of Rejections of the Null
[2]	1. const, bid	3.7520	18.80%
	2. const, log bid	17.5932	97.40%
[5]	1. const, bid	13.5229	90.60%
	2. const, log bid	3.7017	17.80%

Note: The Pearson chi square test is distributed as a chi square with two degrees of freedom under the null hypothesis that the model is adequately specified. The 95% critical value for a chi square with two degrees of freedom is 5.99.

Inconsistent Responses

In additional sets of Monte Carlo simulations the latent WTP amounts were allowed to change between payment questions. These inconsistencies may be the result of the respondent's strategic considerations, confusion, or increased familiarity with the commodity being valued. We draw both latent WTP values from the bivariate versions of the distributions reported in Table 1. This presumes that the changes in WTP amounts are not systematic.

When the correlation between latent WTP variables is moderate (0.5) the double-bounded models are always found to produce biased estimates, whereas the single-bounded models are affected by meaningful bias only when incorrect models are fit. The results are qualitatively similar, but with much larger biases for the double-bounded estimates, when the latent WTP amounts are assumed to be independent. This confirms the finding of Alberini (forthcoming), which shows that the double-bounded estimation procedure is acceptable when the correlation between underlying WTP amounts is sufficiently high.

IV. Concluding Remarks

Based on the results of Monte Carlo simulations, we find that the WTP estimates from dichotomous choice CV survey data are very sensitive to even minor distributional misspecifications and to the assumptions on the shape of the utility function. We conclude that postulating that utility is a linear function of log income is a "high-risk" assumption, in the sense that, should that assumption not be valid, highly biased estimates of welfare change would be obtained. A linear-in-income specification of utility appears to be a somewhat safer choice: having eliminated regressors other than the bid from the probit equation, this specification essentially limits the "damage" to that due to a poor choice of the distribution of willingness to pay.

The the biases resulting from a poor distributional assumption for WTP are, however, not small. We do not find support for our intuition that that should be systematically larger with double-bounded estimation procedures than with single-bounded approaches. In many cases the single-bounded and double-bounded estimated WTP values are not statistically discernible.

Of the statistics and regression diagnostics we look at, only the Pearson chi square test seems to provide hints about the possible inadequacy of a model. Unfortunately, earlier research (Alberini, 1995) shows that this test does not have much power at smaller sample sizes.

Finally, in situations with inconsistencies between responses to the initial and follow-up questions, the double-bounded biases do turn out to be biased, forcing the researcher to rely on the less efficient single-bounded estimates.

We hope that these results, which we view as exploratory rather than final, highlight the importance of specification searches and use of tests of adequacy of the model in empirical work with CV survey data.

References

Alberini, Anna (1995), "Testing Willingness-to-Pay Models of Discrete Choice Contingent Valuation Data," Land Economics, 71, 83-89.

Alberini, Anna (forthcoming), "Bias v. Efficiency of Willingness-to-Pay Estimates: Bivariate and Interval Data Models," *Journal of Environmental Economics and Management.*

Arrow, Kenneth J., et al. (1993), "Contingent Valuation Methodology Report (Report of the NOAA Panel on Contingent Valuation)," *Federal Register*, 15 January 1993, 4602-4614.

Carson, Richard T., Jennifer Wright, Anna Alberini, Nancy Carson and Nicholas Flores (1995), "A Bibliography of Contingent Valuation Studies and Papers," La Jolla, CA, March.

Carson, Richard T., Leanne Wilks and David Imber (forthcoming), "Valuing the Preservation of Australia's Kakadu Conservation Zone," *Oxford Economic Papers*.

Cropper, M.L., L. Deck, N. Kashor and K.E. McConnell (1993), "Valuing Product Attributes Using Single Market Data: A Comparison of Hedonic and Discrete Choice Approaches," *The Review of Economics and Statistics*.

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

Hanemann, W.M. and Barbara J. Kanninen (forthcoming), "Statistical Issues in Contingent Valuation," in Ian Bateman and Ken Willis (eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries* (Oxford: Oxford University Press).

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

Kling, C. (1989), "A Note on the Welfare Effects of Omitting Substitute Prices and Qualities from Travel Cost Models," *Land Economics*, 65, 290-296.

Kling, C. (1991), "Estimating the Precision of Welfare Measure," *Journal of Environmental Economics and Management*, 21, 244-259.

Kopp, Raymond J., and V. Kerry Smith (1993), *Valuing Natural Assets. The Economics of Natural Resource Damage Assessment* (Washington, DC: Resources for the Future).

Lee, Elisa T. (1992), *Statistical Methods for Survival Data Analysis* (New York: Wiley & Sons).

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

Natural Resource Damage Assessment, Inc. (1994), "Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight," Report to the National Oceanic and Atmospheric Administration, La Jolla, CA, September.

Assessing the Content Validity of Contingent Valuation Studies

Richard C. Bishop and Daniel W. McCollum[1]

Abstract: Content validity assessment involves evaluation study procedures. This paper proposes a set of content validity criteria for contingent valuation studies and a rating form for use in assessing how well studies were designed and executed. The form's goal is to help researchers design content valid studies and reviewers to conduct more systematic, balanced validity assessments.

Quoting Mitchell and Carson (p. 190), "The validity of a measure is the degree to which it measures the construct under investigation." In applied welfare economics, the construct is most often one of the Hicksian measures of economic value. Assessing the accuracy of consumer welfare measures is difficult because true Hicksian values are inherently unobservable. Hence estimated values cannot be compared directly with true values to judge the performance of measurement techniques (Bishop et al. 1994). This is the case whether the valuation technique in question is contingent valuation (CV) or one of the methods that attempts to infer values from revealed-preference data. Hence, less direct forms of evidence about the validity of valuation techniques are required.

The raging debate over CV, spawned in part by work surrounding the Exxon Valdez oil spill (Carson et al. 1992; Hausman 1993), is a debate over the validity of the method. Though encouraged by the adversarial context of natural resource damage assessment, this debate is

[1] Professor, Department of Agricultural Economics, University of Wisconsin--Madison, and Economist, USDA, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, respectively. Research supported by the Rocky Mountain Forest and Range Experiment Station and the College of Agricultural and Life Sciences at the University of Wisconsin--Madison. Thomas C. Brown, Patricia A. Champ, Robert K. Davis, John Loomis, and Norman Meade provided many helpful comments on earlier drafts. This paper also benefited greatly from comments of the members of W-133.

symptomatic of a more serious and fundamental gap in applied welfare economics. Advocates of CV are proposing that survey evidence about economic values be accepted in an arena where revealed preference evidence has long dominated. Progress in considering this proposal is hampered by a lack of consensus among economists regarding criteria for judging the validity of welfare estimates, and this is true whether revealed preference measures or CV measures are being considered. More succinctly stated, applied welfare economics lacks a theory of measurement. The goal of this paper is to work toward such a theory. While we focus on CV, we believe that our work has implications for economic measurement more generally. In point of fact, revealed preference measures deserve much more careful and systematic scrutiny than they have received in the past and consistent criteria for revealed preference and CV approaches--and other possible measurement tools--should be a long term goal.

Lacking an economic theory of measurement, CV researchers (e.g., Mitchell and Carson; Bishop et al. 1994, 1995) are turning to other disciplines that have struggled to assess the validity of empirical measures of unobservable constructs, particularly psychology. In psychometrics, the validity of a measure may relate to its content validity, construct validity, or criterion validity (American Psychological Association; Sundberg; Zeller and Carmine; Bohrnstedt). Each of these approaches "offers a different strategy for assessing the measure-construct relationship, and each is applicable to contingent valuation in one way or another." (Mitchell and Carson, p. 190) This paper focuses on content validity.

Content validity assessment involves evaluation of study design and execution.[2] Partly,

[2] Our definition of content validity assessment is significantly broader than that of Mitchell and Carson (pp. 190-192). They focus exclusively on examination of the survey instrument, while we would include all aspects of study design and execution.

it is guided by theory. Measured values will ultimately be interpreted as estimates of values as defined in theory. From a practical standpoint, this means that CV instruments materials must be designed in ways that would support revelation of true values by the consumer of economic theory. Hence, for a CV study to be fully content valid, respondents must have incentives for true value revelation and enough information to make utility maximizing choices. Content validity assessment also asks whether CV study procedures were designed to interact effectively with potential respondents. It is not hard to imagine a study that is strongly linked to theory, yet fails to deal well with real people. Through experience, CV researchers have learned this is not a trivial problem. Finally, to be content valid, the survey, subsequent analysis, and presentation of results must be adequately executed. Here, attention is focused upon such topics as sampling, response rates, and econometric procedures.

Though most did not frame their work in terms of content validity, many writers on CV have addressed issues of study design and execution. The reference operating conditions of Cummings, Brookshire, and Schulze represented an early attempt to explicitly state some validity criteria for CV studies, including content validity criteria. We draw much from Mitchell and Carson, particularly in the area of CV survey design. Recent contributions to the literature on procedural issues include the report of the NOAA Panel on Contingent Valuation (U.S. Department of Commerce) and the recent paper by Hanemann.

Figure 1 will serve as the centerpiece for the paper. We propose it as a tool for systematically rating the content validity of CV studies. The rating form is our attempt to synthesize past literature that relates, in one way or another, to CV content validity and our own experience as researchers. Our purpose is <u>not</u> to attempt to set up ourselves or anyone else as ultimate authorities on the content validity of CV studies. Rather we hope simply to make content validity assessment more systematic and explicit. All of us involved in CV research consider some studies to be stronger than others. Such judgements are based in part on how well such studies appear to have been designed and executed. The rating form is intended to help reviewers, in their various roles[3], to be methodical in evaluating CV study procedures and clearer about their reasons for judging those procedures to be strong or weak. Hopefully, Figure 1 will also help those who conduct CV studies to improve procedures by explicitly stating a set of standards that others will likely use to judge what they have done. The rating form can be viewed as a checklist of considerations that should be addressed in designing and executing studies that aspire to high content validity.

Figure 1 begins with 12 questions about the detailed study procedures. Points are to be assigned to the study under review depending on how well it did in addressing the issues raised under each question. After devoting some attention to preliminary matters in Section I of the paper, each of the detailed questions will be explained and justified in Section II. Following the 12 detailed questions, the rating form asks reviewers to summarize their evaluations of study procedures by adding up the points on the individual items, by explicitly stating any concerns they have that were not covered by the detailed questions, and by rating overall study procedures on a five item scale ranging from excellent to unacceptable. Issues relating to the adding up of points and the overall rating are dealt with in the third section. At the end, the major points of the paper are summarized and some final thoughts offered.

Obviously, judgements about the validity of any study will depend on more than its

[3] That is, those who, as journal reviewers, consultants to decision makers, expert witnesses, or in other such roles, are called upon to evaluate the merits of CV studies.

content validity. Important additional evidence will come from subjecting study results to hypothesis testing based on theoretical expectations (i.e., construct validity testing). How CV has fared in laboratory and field experiments and other efforts to test its criterion validity will also be relevant. Such broader issues are beyond the scope of content validity assessment and are dealt with elsewhere (Mitchell and Carson; Bishop et al. 1994, 1995).

I. PRELIMINARY ISSUES

Professional Judgment and the Burden of Proof

As a point of departure, let us suppose that a proposed "intervention" in the economy affects environmental attributes relevant to some human population. Such an intervention could take the form of a public project, an alteration in environmental regulations, or a new policy that somehow affects the environment. The intervention could also take the form of an accidental or intentional environmental insult such as an oil spill or emission of air pollutants. Suppose further that a CV study has been conducted to estimate the values that members of the affected human population place on enjoying the positive effects of the intervention or avoiding its negative effects. Content validity of such a study would be conducted in the context of two over-arching principles.

First, content validity assessment is inherently a matter of professional judgement. Because there is less than complete consensus about CV procedures, study designers and reviewers must inevitably fall back on personal judgement. At least for the time being, whether procedures are flawed and the seriousness of any flaws remains a matter for individual reviewers to judge based on their interpretations of their own work, if any, and the larger literature. It follows immediately that the cogency of the conclusions from a content validity assessment depend directly on the credentials of the reviewer.

Second, the burden of proof regarding the validity of study's procedures rests with the researchers who designed and executed the study. Replication is problematical in survey research. Furthermore, CV procedures are far from standardized. As a result, content validity assessment involves an evaluation of the detailed study procedures. Researchers must make the case for the content validity of their studies.

Criteria and Points

As we envision it, content validity assessment will require reviewers to consider how well or poorly the study did in addressing a list of procedural issues. Our rating form (Figure 1) is designed to capture the major issues. We do not expect the issues raised there to be particularly controversial. However, we propose that reviewers answer each of the 12 detailed questions with a numerical scores. The maximum attainable numerical scores we assigned to the different dimensions are likely to be more controversial. The current state of the art in CV leaves a great deal of room for debate on the relative importance of different aspects of study design and execution. To deal with this problem, Figure 1 is amenable to whatever weights a particular researcher or reviewer deems appropriate.

As the review of any particular study proceeds, potential flaws in procedures will almost certainly be identified. Such potential flaws will often not be judged fatal, though the possibility of fatal flaws exists and will be dealt with later. Content validity assessment often involves the identification of potential flaws. That is, in the course of the assessment, doubts arise about whether procedures followed might have led to biased results. In more colloquial terms, content validity assessment involves a search for what are commonly termed "red flags." The more such red flags pop up during evaluation of a study, the less valid it will be judged to be. Our scoring system is designed to, in a sense, count red flags, or rather the lack of them.

On any particular item in the form, some studies may easily receive full credit simply because an issue did not arise in that particular case. Other studies may lose points for having neglected to one degree or another the issue or issues highlighted in the question. Under particularly difficult circumstances, a study may receive a low score despite competent efforts to overcome a particularly knotty problem. This would simply reflect the difficult circumstances that are present in that particular case. It should be more difficult to establish the content validity of CV studies in some situations than in others.

Flaws may creep into CV studies through simple lack of foresight on the part of study designers. Furthermore, some flaws are knowingly accepted as compromises required to achieve other goals. For example, in some situations, a referendum format[4] or some other mechanism with theoretically strong incentive characteristics may be very implausible to potential respondents. One might adopt a donation payment vehicle[5] in such situations, notwithstanding its theoretical inferiority. Despite the fact that the researcher made this compromise intentionally and after full consideration of the alternatives, the use of an incentive incompatible mechanism would reduce the content validity of the study (in our opinion!) and this should be recognized

[4] A referendum format frames the CV question in terms a voting for or against the intervention given that an affirmative vote will require some sort of payment.

[5] Framing a CV question around a donation vehicle means that subjects are asked about amounts they would donate toward implementation of the intervention if it is positive or toward avoiding it if adverse consequences predominate.

in the score assigned under the question that relates to incentive compatibility (Question 5, as discussed below).

II. THE DETAILED QUESTIONS

Having laid a foundation for the rating form, we now look at its detailed questions, Question 1 through 12 in Figure 1. In each case, we explain the nature of the issues raised, attempt to assess their importance, and suggest the number of points that we believe that particular question warrants.

(1) Was the theoretical true value clearly and correctly defined?

Study designers may strengthen the link between theory and the CV exercise--thus enhancing content validity--by carefully defining, in theoretical terms, what is to be measured. The simplest model of the consumer's choice problem where environmental quality matters will illustrate. Such a consumer would solve the problem:

$$\max U(X;Q) \text{ subject to } P'X \leq Y,$$

where X is a vector of conventional goods and services that can be purchased at exogenously determined prices P, Q is an exogenously determined vector conveying the status of environmental attributes affecting consumer welfare, Y is income, and U(.) is a "well-behaved" utility function. Assume that the only effect of the intervention in question is to alter the status of environmental attributes, let us say from Q' to Q".

Theory tells us that the maximum level of utility, arrived at by solving the choice problem just stated, can be expressed as an indirect utility function, V(P,Q,Y). Assuming that the

Hicksian compensating welfare measure is relevant, the "theoretical true value" of this intervention to the consumer, which we shall symbolize by T, is defined by

$$V(P,Q',Y) = V(P,Q'',Y-T).$$

Now suppose a CV study is to be conducted to estimate the mean value of T for the policy-relevant population. The benefits of formally considering the theoretical true value are many, as even this simple model illustrates. For example, for respondents to arrive at their estimates of T, they would have to be "well informed" about how the intervention would affect relevant parameters of their choice problem. Respondents would not be well informed if information is unavailable to them that a theoretical consumer would find relevant in solving the utility maximization problem.

Definitions of value should not only be clear, they should be "correct." That is, the researcher should make the theory fit the problem at hand. Some studies will be able to focus on effects of the intervention on environmental attributes alone, as we did in the model just presented. Other studies may have to deal with effects on prices, incomes, and other parameters as well. The timing of both effects and payments may affect true values. Where uncertainty of one kind or another is a potentially significant factor in the theoretical consumer's valuation problem. Designers of CV studies should carefully consider the definition of T applicable in their particular case. Formal theoretical modeling of the valuation problem never hurts. Writing out the equations may seem mundane, but can prove helpful in identifying gaps and flaws in the information and context that will ultimately be provided in the CV scenario.[6] Clearly defining the theoretical true value appropriate in the particular application may help to successfully

[6] In CV jargon, the "scenario" is the part of the survey instrument that communicates to respondents what is to be valued and under what circumstances.

address issues under many of the later questions on the form, especially Questions 3, 4, 5, 11, and 12.

The rating form allows up to 5 points to be assigned to a study depending on how well it defined the true value or values it sought to measure.

(2) Were the environmental attributes relevant to potential subjects fully identified?

In the abstract world of theory, the environmental attributes affecting consumer welfare can be represented by including the vector Q in the direct and indirect utility functions. However, theory alone offers limited guidance regarding which actual attributes are relevant to real world study subjects and which are not. From the potentially large set of attributes of the environment that might be relevant in theory, a subset that human respondents believe affects their welfare must be defined.

Introspection and casual observation on the part of the researchers help to formulate working hypotheses about which attributes might be relevant. For example, it seems likely that attributes affecting human health are important to people. However, once such obviously relevant attributes are identified, it may be necessary to use more formal, empirical methods to sort out which attributes matter. CV studies often employ focus groups for this purpose. Researchers may also observe one-on-one interviews with subjects from the pool of potential respondents. Such interviews and particularly debriefing session with subjects afterwards can help sort out the relevant attributes. Verbal protocols (Schkade and Payne) may be analyzed to further explore how respondents view the attributes. Such "qualitative research techniques," if competently applied, will enhance content validity.

We have allocated up to 10 points for this aspect. How many points to assign to a study will vary depending on the particular circumstances. Studies where respondent-relevant attributes are rather simple and obvious may earn the full 10 points after little or no qualitative research. Other interventions may have effects which are complex and less obviously relevant to people. In such cases, reviewers might assign fewer than 10 points in recognition of the inherent difficulty of the problem.

<u>(3) Were the potential effects of the intervention on environmental attributes and other economic parameters adequately documented and communicated?</u>

Following determination of the environmental attributes relevant to potential study subjects, the next step in study design is to document how the intervention will affect those attributes. This is normally done by finding out what physical and biological scientists know (and do not know) about the effects of the intervention. Impacts on non-environmental parameters such as prices and incomes also need to be documented in cases where they could occur. The more thoroughly such effects were investigated and documented, the higher should be the score on this item.

Once potential effects of the intervention are documented, an instrument to communicate them to respondents must be designed. Real world respondents may come to CV exercises with a great deal of information or no knowledge at all regarding the relevant attributes of the environment.[7] How much knowledge they have prior to the survey must be considered and

[7] There is an ongoing debate among environmental economists about whether the status of an attribute can be "relevant" to consumers who are not aware of it. For one view that has found its way into print, see Bishop and Welsh. Basically, that paper argues that, as a practical matter, real world consumers can not be expected to have full knowledge about all the things affecting their welfare. Obscure and even unknown environmental resources could have value to them.

perhaps assessed in advance through qualitative research. For respondents to be well informed, the knowledge they bring to the CV exercise may need to be augmented with information provided in the scenario.

All else equal, the communication burden placed on the CV scenario will likely be less when respondents have experience-based prior knowledge, than when their prior knowledge was based on media accounts and hearsay. Accordingly, studies that can build their scenarios on experiential knowledge will have the easiest time establishing their content validity. Those that must start from a very limited or non-existent knowledge base will have the most difficult cases to make.

In recognition of the importance of this aspect, the rating form allows up to 10 points to be assigned depending on how well the study documented and communicated the potential effects of the intervention.

<u>(4) Were respondents aware of their budget constraints and of the existence and status of environmental and other substitutes?</u>

Because true values are defined in a framework involving budget-constrained utility maximization, many, including the NOAA Panel, argue that study subjects ought to be explicitly reminded of their budget constraints. Failure to do so would reduce the content validity of a study in the eyes of many potential reviewers.

Thus far, only the elements of the vector Q that would be affected by the intervention have been considered. Theory tells us that the value of environmental amenities affected by the intervention may depend on the status of other amenities that are substitutes for the potentially affected ones. Content validity may, therefore, be enhanced by assessing respondents' knowledge of the existence and status of substitutes during qualitative research and, if necessary, adding information about substitutes to the scenario. Furthermore, the range of substitutes may extend beyond environmental substitutes and include other public and private goods. Presumably complements should also be considered, but there is less emphasis on them in the thinking of many scholars, including members of the NOAA Panel.[8]

Figure 1 recommends up to 5 points be awarded, depending on the reviewer's judgement as to whether subjects were cognizant of their budget constraints and well informed about substitutes.

(5) Was the context for valuation fully specified and incentive compatible?

In addition to providing respondents with needed information about the effects of the intervention, a CV scenario will normally provide them with what we shall term the "context for valuation." Context refers to all dimensions of the proposed transaction dealing in one way or another with the how decisions about the intervention will be made and how money referred to in the CV question will be transferred. Whether the money will be paid to or received by respondents needs to have been clearly spelled out. Points might be lost, for example, if the nature of the value to be expressed was vague (e.g., asking "What is it worth to you?"). Whether the value is to be that of the individual or of the household needs to be clearly stated. Who else will be paying or receiving payment (the so-called "extent of the market," see Smith) may matter for environmental amenities with public goods characteristics. Certainly, theory

[8] The extent to which it is necessary to explicitly deal with budget issues and substitutes in CV scenarios remains a subject for further research. At least one published study (Loomis, Gonzalez-Caban, and Gregory) has found statistically indistinguishable results whether budget constraints and substitutes were mentioned or not.

dictates that the timing of payments has relevance to valuation. A valid CV study will strive to make the context of valuation as complete as possible.

Furthermore, theory raises some rather stern warnings about the incentive properties of CV scenarios. Incentive compatibility of payment mechanisms is an issue even for amenities, such as recreational opportunities, with private-good characteristics. It is well known, for example, that sealed-bid auctions create incentives to bid less than one's maximum willingness to pay, whereas a Vickery auction should lead to full value revelation, all else equal. This theoretical result may have practical relevance to studies using an open-ended CV format.

Where environmental amenities take on public-good characteristics, incentive issues are magnified because of the possibility of free riding and strategic responses. The theoretical strength of the referendum format in this context are widely accepted (e.g., Mitchell and Carson and Hoehn and Randall) and led the NOAA Panel to advocate heavy reliance on referenda in CV studies for purposes of damage assessment. In such circumstances, use of referendum formats, as opposed to voluntary donations, for example, would enhance content validity in the eyes of many reviewers. In our weighting scheme, if the context for valuation is complete and fully incentive compatible, it would be awarded 10 points. Studies with incomplete contexts would fare less well. Fewer points would also be assigned to studies with scenarios that are incentive incompatible in recognition of the potential confusion or strategic responses that such scenarios might induce.

(6) Did survey participants accept the scenario? Did they believe the scenario?

CV researchers and others (e.g., the NOAA Panel) have come to recognize that it is important that the scenario not only be communicated effectively, but that respondents accept it. A study subject accepts the scenario when he or she implicitly agrees to proceed with the valuation exercise based on the information and context provided. Scenario rejection can lead either to poor quality valuation data or item non-response for CV questions.

Content validity would be enhanced if respondents not only accept the scenario, but believe it. Those writing on CV often emphasize that it involves "hypothetical" valuation, but some scenarios are more hypothetical than others. In many settings, asking study subjects to play "what if" games in order to value the intervention is unavoidable because a fully believable scenario is impossible to construct. However, in some circumstances, it may be possible to construct a scenario with a high degree of plausibility.

An example from the author's current research will illustrate. The work focuses on possible modifications in how Glen Canyon Dam on the Colorado River is operated. Changes may be needed to protect and enhance resources downstream in the Grand Canyon. Modifying dam operations would reduce its ability to generate electricity on-peak. A very likely result will be increases in how much some households in several western states will pay for electricity. One sampling frame for the CV study on this problem is the potentially affected electricity consumers. A referendum format is being used and the payment vehicle for this sampling frame will be electricity costs to these households. Focus groups showed that subjects found it very plausible that they would have to pay more for electricity if dam operations are modified. This enhances the credibility of their responses.

The rating form suggests that reviewers assign up to 10 points depending on their evaluation of whether respondents accepted the scenario and whether they found it believable.

To earn all 10 points a study would have to demonstrate rather unambiguously that respondents both accepted and believed the scenario. We would personally assign a fairly high score, perhaps 7 or 8, to a study that was very forthright about the hypothetical nature of the valuation exercise (thus foreclosing belief), but showed clear evidence that respondents nevertheless accepted the scenario. Whether respondents accepted and believed the scenario is admittedly difficult to determine, but some evidence can often be mustered. After careful consideration of the instrument, reviewers will no doubt form judgements about the plausibility of the scenario and the potential for scenario rejection. Furthermore, whether potential respondents accept and believe the scenario can be intentionally evaluated during focus groups and other procedures followed during qualitative phases of the research. Reports of such activities may help to reviewers evaluate these two dimensions. Furthermore, debriefing questions may be included in the survey to help determine rates of acceptance and belief.

(7) How adequate and complete were survey questions other than those designed to elicit values?

CV surveys typically include many questions other than those intended to elicit values. Several different objectives may be involved. For one, CV researchers often find it desirable to investigate respondents' motives for answering CV questions as they did. The exact form of such questions depends on both the form of the CV question and the researcher's judgement. For example, open-ended CV questions are often followed by questions designed to tell explore what respondents intended when they responded with a zero. A respondent may actually have had a zero value for the intervention, but a zero may also have been intended to communicate that the respondent did not know her value, refused to place values on the intervention, rejected the scenario, or hoped that her response would reduce fees actually paid. The NOAA Panel, which, as noted already, recommended that a referendum format be used, also recommended that voting be followed by a question in an open-ended format asking respondents to explain why they voted as they did.

Additional questions may be included in the survey to provide evidence of its content validity. For example, appropriately worded questions could help evaluate whether respondents understood descriptive material in the scenario. Many past studies have included follow-up questions to attempt to identify strategic responses.

Other questions may also be included to assess the construct validity of the study. Construct validity tests normally involves hypotheses about relationships between answers to CV questions and other variables either in cross tabulations or in multiple-regression analyses (Bishop et al. 1994). Many types of questions can be included in the survey to support such analyses. For example, the NOAA Panel recommended cross tabulations of valuation responses with income, knowledge of the site, prior interest in the site for visitation or other reasons, environmental attitudes, attitudes toward big business, distance of residence from the site, understanding of the valuation task, and willingness and/or ability to perform the task.

Such survey questions need to be scrutinized as part of content validity assessment. Only if they are well designed will responses provide supporting data needed to meet the various objectives just noted. Because such questions are so important for construct validity testing and other purposes, the rating form assigns 10 points to this dimension.

(8) Was the survey mode appropriate?

Mail surveys are attractive to CV researchers because they are the least expensive of the major modes. There also may be methodological reasons for choosing a mail approach. Mail is preferred by some researchers because mail instruments give them complete control over the information and context communicated to potential respondents. Other researchers shy away from mail surveys because of limited reading skills of potential respondents from the general population, even in the US and other countries where literacy rates are relatively high. Furthermore, even the more literate respondents may be reluctant to try to read and digest large amounts of written material about the intervention and its consequences.

Telephone interviews are more expensive than mail surveys and are limited in the amount of information and context that can be communicated during a brief phone call. Effective communication may require presenting respondents with visual aids such as charts, graphs, and photographs. This will not be feasible in a survey conducted entirely by phone. On the other hand, it is somewhat easier to get reasonably high response rates by phone than by mail and reading skills are not involved.[9]

Personal interviews can make communication easier because of the personal contact between respondent and interviewer. More information can normally be provided than would be possible by mail or over the phone. Conducting surveys in person may increase response rates. However, in-persons surveys with high response rates are very expensive. Likewise, the presence of an interviewer may influence responses.

---

[9] Some researchers believe that use of the telephone for solicitation of sales, donations, and political support, sometimes in the guise of surveys, may be eroding the effectiveness of the telephone mode for actual surveys.

From the perspective of content validity assessment, survey mode must be appropriate for the study goals and the complexity of the information and context that need to be communicated. If the goal is to value a recreational experience that is quite familiar to respondents, for example, then a mail survey may be quite adequate. If the goal is to estimate non-use values for a spill that had complex impacts on environments unfamiliar to respondents, then, as the NOAA Panel recommended, personal interviews would appear to have a large advantage. Using a mail or telephone survey in such a situation would be grounds for questioning the content validity of a study. This is not to say that a mail or telephone survey would necessarily be ruled out. However, in the eyes of many CV researchers, an extra burden of proof would rest on the study team to provide evidence that the mail or telephone procedures worked well.

Many CV researchers stress the importance of survey mode and we agree by assigning up to 10 points to this item.

(9) Were qualitative research procedures, pretests, and pilots sufficient to find and remedy identifiable flaws in the instrument and associated materials?

Once survey designers have roughed out an instrument and related documents based on their understanding of how respondents will react, qualitative research is often needed to refine the instrument.[10] For example, focus group participants may be asked to complete a draft mail survey and then discuss it with the group leader. Or, an instrument designed for personal interviews can be tested in observed interviews. During such interviews, and afterwards in debriefing sessions with the subjects, researchers can try to identify ways that the instrument is being misinterpreted or if information provided is incomplete or otherwise inadequate. Possible

[10] Circulating the instrument to knowledgeable colleagues for review may also be helpful.

improvements can be tested as well. Qualitative testing should not only involve verbal materials but also any photographs or other visual aids.

Formal pretesting and piloting[11] of a nearly finished instrument may also improve it. Statistical analyses of responses provides a preview of what to expect in the final results and can help diagnose problems. Interviewers often help to identify places where in-person and telephone questionnaires can be improved. Interviewers can also be instructed to record verbatim any remarks by respondents about the survey questions and information presented. Though less effective, subjects in mail pretests and pilots can be asked to write comments in the margins. A subsample can be contacted by telephone to probe for flaws in a draft mail instrument. Through such procedures, the study design can be tested under field conditions, enhancing content validity in the process.

Though we share the now commonly accepted view that qualitative research can be invaluable in the design of CV surveys, its limitations in supporting validity must also be recognized. The typical study report will include only a terse statement such as, "Four focus groups were conducted." Little or nothing is said about the extent to which the focus groups succeeded in working the "bugs" out of the instrument and associated documents. Standard procedures for applying qualitative research tools and reporting the results do not exist, or at least have not found their way into everyday practice in economics. This may be a fruitful area for research. In the meantime, reviewers of CV studies may have to take the "quality" of

[11] Pretests are distinguished from pilots by their small and more convenient samples. The goal of pretests is to identify major problems with the instrument and survey execution procedures that will be apparent even for small samples. Question wording that will confuse large numbers of respondents or lead to large item non-response may become apparent, for example. Pilot studies are conducted to further refine question and information wording, test proposed procedures for the final survey under field conditions, and investigate the likely statistical properties of final results.

qualitative work more or less at face value. An exception may be litigation, where details about procedures and results can be ferreted out from audio and video records, from written reports entered into evidence, and from depositions and cross examination.

CV can be applied in such diverse settings that generalizations are not possible regarding how much qualitative research, pretesting, and piloting are needed in any particular case. At one extreme are studies of relatively straightforward interventions, where there is a long history of past research upon which to draw. In such cases, instruments may require little preliminary testing. At the other extreme are non-use studies involving environmental resources unfamiliar to large numbers of potential respondents. Hence, judgements about the appropriate amount of preliminary work must take specific circumstances into account. Up to 5 points are to be assigned to this aspect under our version of the rating form.

(10) Given study objectives, how adequate were procedures employed to choose study subjects, assign them to treatments (if applicable), and encourage high response rates?

Adequate population definition, sampling, and survey procedures depend on study objectives. To allow for this fact, we will distinguish between two different kinds of studies. Some studies involve exclusively methodological goals. One might, for example, design a study to compare the results of open-ended CV questions with those from a bidding game for the same amenity. Other studies have as a major goal the estimation of values for a population of individuals, either in the context of policy analysis or litigation. For convenience, we will term the former "methodological studies" and the latter "applied studies." Applied studies may also have methodological goals. Their distinguishing feature is that they have the ultimate goal generalizing results from sample to population.

For methodological studies, procedures for choosing subjects and allocating them among treatments are mostly a matter of common sense. Where new CV procedures or hypotheses about CV data are to be tested, one would hope to eventually conclude something about how CV would perform in applied studies under normal circumstances. Hence, one might not want to choose kindergartners as subjects. Content validity might suffer a bit if only undergraduates were used as subjects since their responses might be very different from general population samples used in many CV studies. However, at the other extreme, fastidious sampling from the general population or some other group would normally not be required for methodological studies. If the goals of the research are purely methodological, the self-selection bias inherent, for example, in recruiting from the general population subjects who are willing to come to a laboratory and participate in an experiment would probably not be a large red flag in most researchers' and reviewers' judgement. In studies involving multiple treatments, assignments to cells should, of course, be random. In field (as opposed to laboratory) studies, follow-up procedures to increase response rates could normally be less rigorous than in an applied study. In sum, the validity of implementation steps for methodological studies focus mainly on the reasonableness of the procedures in light of the study goals.[12]

Applied studies, on the other hand, must satisfy more rigorous standards as far as sampling and response rates are concerned. Either random or stratified random samples are required which will support extrapolation of value estimates from sample to population. Furthermore, potential non-response bias must be addressed. The best way to head off non-

---

[12] This relaxed attitude toward methodological studies does not carry over to most other aspects of study design. In fact, one might argue that, in most respects, the requirements for design of methodological studies should be even more rigorous than for applied studies.

response bias is by gaining a high response rate in the first place. Survey researchers have well developed procedures for doing so. Various methods to gain a rough idea of the potential seriousness of non-response bias are available. An example would be to compare reported socioeconomic characteristics of respondents with published statistics for their Census tracts. In some cases, population statistics are available in sufficient detail to allow weighting of the sample to represent the population. Careful attention to this issue enhances content validity.

Up to 10 points can be allocated to a study depending on how well it dealt with sampling, non-response, and related details within the context of its overall objectives.

(11) Was the econometric analysis adequate?

Once the responses are in, high content validity requires that the data be competently coded and entered into computer files for analysis. Success here again is simply a matter of using common sense. For example, verification of data is often facilitated by entering it twice and reconciling the data files.

The analysis itself should employ econometric procedures that are appropriate to the data and the inferences that are to be drawn. Economists are normally well trained in this area. Assessing this aspect of content validity is mostly a matter of verifying that analysts have employed their tools properly. We assign 10 possible points to this aspect.

(12) How adequate are the written materials from the study?

The final step in study execution involves reporting study design and execution procedures and study results. Needs here will vary depending on study goals and the expected audience for the report. A journal article might stress technical and methodological details, while a report for policy makers might stress final results and policy implications. Study reports should reflect such objectives.

Content validity assessment itself requires rather complete reporting. Because the burden of proof for content validity rests with the researchers, studies that do not provide thorough and complete reports can not be presumed to have high content validity. This no doubt was part of the motivation for the NOAA Panel's rather severe requirements for reports:

> Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties . . .

From the somewhat broader perspective taken in this paper, the ideal study report would also include a clear statement of the study goals, a definition of the true value to be estimated, a description of the intervention and its effects on environmental amenities, and a fairly detailed summary of the procedures followed throughout the study.

The rating form asks reviewers to assign up to 5 points for this aspects.

III. OVERALL EVALUATION QUESTIONS

(13) Total Points

Once the detailed study procedures have been scored, the rating form suggests that the reviewer add up the points. Some reviewers may wish to skip this step, arguing that it implies a degree of quantitative precision far beyond what can be hoped for under the current state of the art in CV. We can certainly appreciate the reasons for such a reservation. We would nevertheless encourage reviewers to struggle with the numbers, including their aggregate value. We believe that doing so will promote balance in appraisals of content validity. In considering such a complex set of issues, one may tend to focus too much attention on some aspect that seems particularly well done or innovative, or on some flaw that is particularly glaring. Without the discipline imposed by assigning and summing the numbers, too little weight may implicitly be assigned to other study procedures that were done well or poorly. Struggling with the numbers and aggregating them will help avoid such imbalances. Furthermore, it may encourage deeper consideration of the criteria themselves. Particularly after several applications of the rating form, one may feel that the score for a given study seems too high or too low. If so, this may indicate that the weights on the individual items are not in keeping with that reviewer's more fundamental judgements about the relative importance of the various issues raised in the individual detailed questions. The weights may need to be adjusted. In the process of considering this issue, reviewers can force themselves to more carefully consider the criteria they apply and the relative importance they place on different criteria.

(14) Are there other concerns relating to the design and execution of the study that have not already been addressed?

At this point, before the final step in the rating process, we confront two problems. First, CV study procedures still involve many dimensions about which widely-respected researchers disagree. There may well be dimensions that some feel are essential that are not even mentioned here. Second, Question 14 will come into play when special circumstances not ordinarily faced in CV studies are present. For example, timing of survey administration may be an issue in some circumstances but not in others. Suppose injuries due to a large oil spill are to be valued. Doing a CV study too soon afterward might be challenged on the grounds that respondents were still in a state of shock and outrage, and answered the survey in ways that reflected emotions of the moment. Resulting value estimates would be of questionable validity because they might not be robust over time.

Question 14 provides the opportunity for reviewers to write in concerns and issues not raised elsewhere in the rating form, including those that were more or less unique to the particular study being reviewed.

The final step in the content validity assessment is to sum up the reviewers overall evaluation of the study by responding to Question 15.

<u>(15) Considering the issues raised in Questions 1 through 12, your total score as calculated for Question 13, and any additional issues raised under Question 14, how would you rate this study overall?</u>

______ Excellent
______ Good
______ Fair
______ Poor
______ Unacceptable (Study Fatally Flawed)

The response to this question should help interpret the numerical scores and particularly the total points. Suppose, for example, that a study received an aggregate score of 50 points. Such a score would surely mean the reviewer had many concerns, but might not be sufficient to convey just how serious those concerns were. A score of 50 would almost certainly be inconsistent with a rating of "excellent" or even "good," but would not convey whether the study was judged "fair," or "poor" or even "unacceptable." The qualitative rating in the final question should help to clarify how serious the potential flaws in the study were judged to be. A rating of "unacceptable" would signify that a study had fatal flaws. This response would be appropriate if the study failed to meet the reviewer's minimum standards under any of the detailed questions in Figure 1 or a combination of questions or if concerns described under Question 14 were particularly compelling. Suppose, for example, that a study employed telephone interviews in a way the reviewer judged to be not at all adequate to provide sound CV data. Such a study would fail to meet this reviewer's minimum requirements under Question 8. The reviewer would declare the study unacceptable under Question 15 regardless of the total points it earned when the detailed question scores were added. A study that failed to communicate well, neglected to provide a minimally adequate context, or failed miserably elsewhere should simply be identified as unacceptable.

The link between study goals and the criteria for fatal flaws is important to remember. A study designed to be a first preliminary investigation of benefits or natural resource damages, for example, should not be held to the same standards as one that is designed to serve as a basis for an important policy analysis or a final damage estimate. A low-budget study designed to serve primarily as a student project might leave many loose ends that would be unacceptable in a study destined to be used to set damage in an important court case.

IV. SUMMARY AND SOME FINAL THOUGHTS

In this paper, we have attempted to clarify and systematize an approach to content validity assessment for CV studies. A content valid CV study is rooted throughout in a clear theoretical definition of the true value of the intervention. At the heart of such a study will be its scenario. Based on well-documented evidence of the respondent-relevant effects of the intervention, a sound scenario effectively communicates the potential effects of the intervention to respondents. It includes whatever information they need regarding substitutes for the environmental resources in question and may need to remind them of their budget constraints. It also includes a fully specified and incentive compatible context for valuation. It does all this in ways that potential respondents will accept and, if possible, believe.

Looking beyond the scenario, a content valid survey instrument will include well-designed questions to support construct validity testing and achieve other goals. The mode chosen for administering the survey will be appropriate to the complexity of the scenario and the ultimate goals of the study. Prior to administration, the instrument will have been subjected to sufficient qualitative investigation, pretesting, and, if needed, piloting to work out as many bugs as possible. Econometric analysis of the results will have been adequately performed and final results effectively reported.

When studies fall short of these ideals, as nearly all will, they may still have substantial merits. Content validity is normally a matter of degree. However, some studies will fall below minimal standards and be judged content invalid. Other studies, though they may not be rejected outright, may still be viewed with substantial reservations because of possible flaws in design and execution.

To admit evidence from surveys into applied welfare studies, where revealed-preference data have historically dominated, would be a big step for economists. Whether contingent values ought to be considered "admissible evidence" should be approached in a cautious, but open-minded, way based on carefully thought out "rules of evidence." Thus do the social sciences progress. Drawing on its sister disciplines, economists can evaluate this new direction based on content, construct, and criterion validity. Content validity deserves more attention if real progress is to be made.

Figure 1

CONTENT VALIDITY RATING FORM FOR CONTINGENT VALUATION STUDIES

(1) Was the theoretical true value clearly and correctly defined? (5 points) . . . . ________

(2) Were the environmental attributes relevant to potential subjects fully identified (10 points). . . . . ________

(3) Were the potential effects of the intervention on environmental attributes and other economic parameters adequately documented and communicated? (10 points) . . . . ________

(4) Were respondents aware of their budget constraints and of the existence and status of environmental and other substitutes? (5 points) . . . . ________

(5) Was the context for valuation fully specified and incentive compatible? (10 points) . . . . ________

(6) Did survey participants accept the scenario? Did they believe the scenario? (10 points) . . . . ________

(7) How adequate and complete were survey questions other than those designed to elicit values? (10 points) . . . . ________

(8) Was the survey mode appropriate? (10 points) . . . . ________

(9) Were qualitative research procedures, pretests, and pilots sufficient to find and remedy identifiable flaws in the instrument and associated materials? (5 points) . . . . ________

(10) Given study objectives, how adequate were procedures employed to choose study subjects, assign them to treatments (if applicable), and encourage high response rates? (10 points) . . . ________

(11) Was the econometric analysis adequate? (10 points) . . . . . ________

(12) How adequate are the written materials from the study? (5 points) . . . . ________

<u>(13) TOTAL POINTS:</u>. . . . . ________

Figure 1 (continued)

(14) Are there other concerns relating to the design and execution of the study that have not already been addressed? ______________________

____

____

____

(15) Considering the issues raised in Questions 1 through 12, your total score as calculated for Question 13, and any additional issues raised under Question 14, how would you rate this study overall?

_____ Excellent
_____ Good
_____ Fair
_____ Poor
_____ Unacceptable (Study Fatally Flawed)

REFERENCES

American Psychological Association. 1974. *Standards for Educational and Psychological Tests*. Washington, D.C.

Bishop, Richard C., Patricia A. Champ, Thomas C. Brown and Daniel W. McCollum. "Non-Use Values: Theory and Empirical Applications." Paper presented at the Werner Reimers Meeting on Determining the Value of Non-Market Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods, Bad Hamburg, Germany, July 24-29, 1994, proceedings forthcoming.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(5):926-930.

Bishop, Richard C., and Michael P. Welsh. 1992. "Existence Values in Benefit-Cost Analysis and Damage Assessment." *Land Economics* 68(4):405-417.

Bishop, Richard C., Michael P. Welsh, and Thomas A. Heberlein. 1993. "Some Experimental Evidence on the Validity of Contingent Valuation." Unpublished paper, Department of Agricultural Economics, University of Wisconsin--Madison.

Bishop, Richard C., Patricia A. Champ, and Daniel J. Mullarkey. 1995. "Contingent Valuation." In *The Handbook of Environmental Economics*, Daniel W. Bromley, Ed. Cambridge: Basil Blackwell Ltd. pp. 629-654.

Bohrnstedt, George W. 1983. "Measurement." In *Handbook of Survey Research*. Peter H. Rossi, James D. Wright, and Andy B. Anderson, Eds. New York: Academic Press. pp.69-121.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill." Report Submitted to the Attorney General of the State of Alaska. Reprinted by Natural Resource Damage Assessment, Inc.

Cummings, Ronald G., David S. Brookshire, and Willam D. Schulze, eds. 1986. *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method*. Totawa, N.J., Rowan and Allanheld.

Hanemann, W. Michael. "Valuing the Environment Through Contingent Valuation," *Journal of Economic Perspectives* 8(4):19-44 (Fall 1994)

Hausman, Jerry A., Ed. *Contingent Valuation: An Assessment*. Amsterdam: North-Holland Publishing, 1993.

Hoehn, John P., and Alan Randall. 1987. "A Satisfactory Benefit Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management* 14(3):226-47.

Loomis, John, Armando Gonzalez-Caban and Robin Gregory. 1994. "Substitutes and Budget Constraints in Contingent Valuation." *Land Economics* 70(4):499-506.

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

Schkade, David A., and John W. Payne. 1994. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation." *Journal of Environmental Economics and Management* 26:88-109.

Smith, V. Kerry. 1993. "Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal." *Land Economics* 69(1):1-26.

Sundberg, N.D. 1977. *Assessment of Persons*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

U.S. Department of Commerce. National Oceanic and Atmospheric Administration. 1993. *Natural Resource Damage Assessments under the Oil Pollution Act of 1990.* Federal Register 58(10):4601-4614.

Zeller, Richard A., and Edward G. Carmines. 1980. *Measurement in the Social Sciences: The Link Between Theory and Data*. Cambridge: Cambridge University Press.

IMPROVING VALIDITY EXPERIMENTS OF CONTINGENT VALUATION METHODS: RESULTS OF EFFORTS TO REDUCE THE DISPARITY OF HYPOTHETICAL AND ACTUAL WTP

John Loomis,[1] Thomas Brown,[2] Beatrice Lucero[2] and George Peterson[2]

March 1995

Acknowledgement

We would like to thank Helen Neill, University of Nevada-Las Vegas and Philip Ganderton, University of New Mexico for their suggestions on a draft of our experimental design and for Phil's constructive criticism of an earlier draft of this paper. Richard Bishop, University of Wisconsin, Kevin Boyle, University of Maine, Tom Stevens, University of Massachusetts and Richard Walsh, Colorado State University, all provided valuable suggestions to improve the clarity of the presentation. These helpful individuals are in no way responsible for what follows.

1. Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO

2. Rocky Mountain Forest and Range Experiment Station, USDA Forest Service, Fort Collins, CO

IMPROVING VALIDITY EXPERIMENTS OF CONTINGENT VALUATION METHODS: RESULTS OF EFFORTS TO REDUCE THE DISPARITY OF HYPOTHETICAL & ACTUAL WILLINGNESS TO PAY

ABSTRACT

Independent samples and paired responses of adults were used to test differences between hypothetical and actual cash willingness to pay (WTP) for an art print. One of the treatments used a standard open-ended WTP question. The other attempted to overcome hypothesized reasons for divergences between cash and hypothetical WTP by requesting that respondents not report what they thought a fair price was for the good, but rather to act as if this was a real market and to take their budgets into consideration. The results suggest rejecting the equality of hypothetical and actual WTP, but the differences are smaller than in other similar recent experiments. Our open-ended WTP question format resulted in hypothetical WTP that was two to three times larger than actual WTP.

I. TESTING CRITERION VALIDITY OF CONTINGENT VALUATION

One of the long standing criticisms of the contingent valuation (CV) method is that stated willingness to pay (WTP) may be a poor indicator of actual WTP. Although there have been over a thousand applications of CV (Carson et al. 1994), few studies have examined the validity of CV responses, and fewer still have tested criterion validity. Criterion validity tests of CV compare hypothetical WTP to actual cash payments. The scarcity of criterion validity tests may reflect the difficulty in finding or creating a criterion—an actual cash measure to which the hypothetical (contingent) measure is appropriately compared.

Past criterion validity tests of CV have been of two basic types: field experiments and laboratory experiments.[1] The field experiments fall into two groups: those of WTP for hunting permits (Bishop and Heberlein 1979; Welsh 1986), and those involving contributions (essentially donations) towards environmental improvements (Seip and Strand 1992; Duffield and Patterson 1992; Champ et al. 1994). The latter set of studies has the advantage of focusing on public goods, which is an important focus of CV, but are potentially subject to free-riding when measuring actual payments because they use a donation payment vehicle. Both types of field experiments are difficult to arrange and expensive to administer, which has led several researchers to employ laboratory experiments.

Lab experiments comparing actual cash and hypothetical WTP have the advantage of careful control of procedures, use of a well-defined good, and avoidance of free riding. Most lab experiments have used common market goods,[2] such a chocolate bar (Kealy et al. 1988), a house plant (Boyce et al. 1991), a painting or map (Neill et al. 1994), and a juicer, calculator, or box of chocolates (Cummings et al. 1995). Our study has important similarities to the Neill et al. study. Their study included two experiments in which, in the actual payment conditions, student subjects were required to pay "out of pocket" with their own funds, although short-term interest free loans were available. In the first experiment, hypothetical open-ended statements of the maximum WTP for an original painting were compared with another group's real money

[1] One exception to this is the comparison by Brookshire and Coursey (1987) of a field CV with a laboratory use of a Smith auction for valuing changes in tree density of a neighborhood park.

[2] An exception is Coursey et al.'s (1987) Vickery auction study of payment to avoid experiencing a foul-tasting liquid—a private good, but certainly not a typical market good.

bids for the painting elicited in a Vickery (second price) auction. In the hypothetical condition, subjects were asked, "if the painting were to be made available for you to purchase here and now, what is the maximum amount that you would pay for it?" Mean payments were $38 and $9 in the hypothetical and actual payment conditions. In the second experiment, which involved bidding for a framed map, hypothetical payments were nine times actual payments (once bids above $1000 were omitted) and a third condition—using a hypothetical Vickery auction—verified that the difference between hypothetical and actual payments was not attributable to the difference between the open-ended and second price auction elicitation procedures.

The rationale for such studies as indications of the validity of CV (Cummings et al. 1995) is that estimating WTP for a private market good should be *easier* for subjects than estimating WTP for a nonmarket good (i.e., if people cannot estimate what they would pay for an observable private good, then how can they estimate what they would pay for a complex and unfamiliar environmental good?). However, using market goods to test criterion validity may encounter its own unique problems having to do with price cues. First, in stating their WTP in the hypothetical market, some individuals may state what they believe a fair price is or what they guess the market price to be, rather than what they would pay. If some subjects perceive their task as playing the "price is right", this reduces the insight these experiments provide to estimating the value of public goods for which no obvious price exists. Second, some subjects may answer what they would pay *if* they were in the market for such a good, rather than what they would pay *now* if given the opportunity to actually purchase the good. Further, respondents may project into a less constrained environment when answering the hypothetical WTP question as compared to the actual payment question. For example, they might be stating what they would pay after payday. These conditions often do not parallel the cash experimental treatments where respondents are asked to pay, here and now.

Another issue raised by the set of previous criterion validity studies is whether it is essential to estimate hypothetical and actual WTP using independent samples. Most of the studies listed above used independent samples, but the Seip and Strand (1992), Kealy et al. (1988), and Boyce et al. (1991) studies asked the same respondents to participate first in hypothetical and then in actual payment conditions. Were these latter three studies subject to order effects, whereby subjects' hypothetical responses affected their later actual bids? We compared independent and sequential (paired) actual WTP responses to help answer this question.

Our research objectives were to (1) compare estimates of hypothetical and actual WTP for a market good; (2) test whether some of the disparity between hypothetical and actual WTP for a market good can be removed by carefully instructing respondents in the hypothetical treatment to answer the correct question; and (3) compare independent sample responses of actual WTP with sequential/paired responses. To avoid potential problems with using student subjects to estimate payments, we used nonstudents.

II. RESEARCH DESIGN

A laboratory experiment was designed to compare estimates of WTP elicited using an open-ended question format in both hypothetical and real markets for the same good. The experiment had three treatments: (1) *WTP(h:no reminder)*, where hypothetical WTP was asked in a manner similar to a standard CV survey, (2) *WTP(h:reminder)*, where hypothetical WTP was asked after subjects were reminded not to give what they think a fair price is or what the good sells for and to act as if they were in a real market with their real budget, and (3) *WTP(a)*, where actual WTP was requested.

As Table 1 indicates, the experiment used three sessions, each with a separate sample of respondents. Session 1 used treatments 1 and 3, session 2 used treatments 2 and 3, and the last session used only treatment 3. When the WTP(a) treatment is administered after a hypothetical WTP treatment, the resulting WTP measure is designated below as *WTP(a_p)*, where "p" indicates a measure "paired" with a hypothetical measure.

Our first objective, to compare hypothetical and actual WTP, was tested with the following null hypotheses:

1. WTP(h:no reminder) = WTP(a)
2. WTP(h:reminder) = WTP(a)
3. WTP(h:no reminder) = WTP(a_p)
4. WTP(h:reminder) = WTP(a_p).

WTP(a), the independent measure of actual WTP, was considered the criterion for the validity test because it was not potentially contaminated by a hypothetical WTP treatment. Tests of hypotheses 1 and 2 are therefore the principal criterion validity tests. Although the actual WTP responses used in hypotheses 3 and 4 are conditional (i.e., potentially subject of order effects), they do provide comparison in which differences among respondents are controlled.

Our second objective, to see if disparity between hypothetical and actual WTP could be lessened by careful reminders to respondents in the hypothetical treatment to estimate their WTP as if the good were really for sale, was tested with the following null hypothesis:

5. WTP(h:reminder) = WTP(h:no reminder).

We also performed multi-variate tests of the hypothesis that without the reminder subjects report what they think the object sells for rather than their own personal WTP.

Our third objective, to compare separate and paired (sequential) estimates of actual WTP, was tested with the null hypothesis that the hypothetical WTP question would have no effect on subsequent actual WTP. The two paired measures of actual WTP allowed two tests, where p1 refers to the session 1 measure and p2 refers to the session 2 measure:

6. $WTP(a) = WTP(a_{p1})$

7. $WTP(a) = WTP(a_{p2})$.

III. EXPERIMENTAL DESIGN

Subjects

University clerical and administrative staff in academic and non-academic units were recruited and paid $20 for attending a 45 minute session on campus. The sample sizes are given in Table 1 and range from 30 to 35 people per session. The same researcher conducted all of the sessions, following a script that was identical except for treatment effects.

Nature of the Good

The good chosen for the experiment—an art print—reflected several desirable features. First, art prints are infrequently purchased and different prints sell at quite different prices, so people would not be likely to know the market price of a given print. The objective was to minimize the likelihood that respondents would simply state its known market price. Second, the good had readily observable characteristics, so there was minimal ambiguity in terms of what the product was. Finally, the good was not too expensive, so it could be paid for in cash or from the current balance in respondents' checking accounts.

We pretested several wildlife art prints on a separate sample of university staff, and settled on a signed wildlife print of a wolf standing in a forest. Using the wolf print, we conducted five pre-test sessions with different samples of university staff to better understand the thought processes used in both the actual cash and hypothetical market scenarios. Based on wrap-up discussions and written comments elicited during these sessions, several revisions were made to procedures and instructions. Changes included adding questions for respondents to rate the prints prior to the auction, changing the wording in the reminder statement to counter tendencies of the respondents to not fully consider their current budget, and adding phrases to the actual payment treatment making it very clear that the print was really going to be sold. This process continued until respondents' comments indicated they understood the task before them in each treatment as we intended.

Setting of Experiments

All the sessions were held in a classroom with participants sitting at every other seat to maintain privacy and avoid discussion among participants. At the beginning of a session, participants were individually shown the art print. They were then asked to rate, using a five-point Likert scale, how well they liked the print, whether they would buy it for themselves or a friend, and if they were in the market for art prints. Next, individuals were instructed to read and complete a bid submittal page which varied depending on the session (with sessions 1 and 2 receiving a hypothetical treatment and session 3 receiving the actual payment treatment, as shown in Table 1). When everyone had finished, the sheets were passed forward. All respondents then filled out a sheet on their demographics. Respondents in sessions 1 and 2 were then given the actual cash treatment. In all three sessions, following the actual cash treatment, the winner was announced and asked to come forward to complete his or her purchase in front of the group, but the winning price was not announced. Individuals were allowed to pay with cash or check or sign a promissory note payable within three weeks.

Wording of WTP Questions

In session 1, the wording of the WTP(h:no reminder) question was:

> *You are being asked to participate in a hypothetical sealed bid auction for this print. We would like to know the maximum amount of money you would pay to take this art print with you at the end of this session, if this one art print were actually for sale, and you would have to pay by August 19, 1994.*
>
> *Now please write down the maximum dollar amount you would be prepared to pay for this art print. I would bid $_________.*

This wording was patterned after Neill et al. (1994).

In session 2, the wording of the WTP(h:with reminder) question was:

> *You are being asked to participate in a hypothetical sealed bid auction for this print. We would like to know the maximum amount of money you would pay to take this art print with you at the end of this session, if this one art print were actually for sale.*
>
> *At this time in the survey, we are NOT asking what you think the art print might sell for in a store or what you think its fair price is. Rather, we want to know the maximum amount of money that you would honestly be prepared to pay right now to buy the art print you are being shown if you would really be required to pay your bid amount with cash, write a check today, or sign a Promissory Note payable on or before August 19, 1994. Please take into consideration your budget and what you can afford to pay. If what you would pay is different from what you judge a fair price to be, that is OK. We want to know what you would actually be prepared to pay for the art print.*
>
> *Take a few moments to think about what you honestly would be prepared to pay for this art print if it were being offered for sale to you today and it would go to the highest bidder. Although the question is hypothetical, we want you to answer as if it were for real—as if you were participating in a real sealed-bid auction and would really have to pay your dollar amount if you were the highest bidder.*

Participants in this hypothetical treatment with reminder were asked to read the foregoing instructions and then stop and wait for further instructions. When it was apparent that everyone had finished reading the instructions, the interviewer then reiterated the foregoing instructions verbally to the participants. They were then asked to proceed. The questionnaire proceeded as follows:

> *Now, please write down the maximum dollar amount you would be prepared to pay for this art print. I would bid, and would really be prepared to pay, $________.*

As can be seen, the reminder attempted to more fully place the individual in the frame of mind of a real market situation, without actually requiring them to pay. This statement was developed after discussions with pre-test participants indicated that they were in a different frame of mind when answering the hypothetical WTP questions as compared to a follow-up actual cash question. Second, we wanted individuals to report their WTP for the print rather than attempt to estimate what they thought a reasonable price would be in a store.

Wording of the actual cash WTP question was:

> *As part of this experiment, we are now going to conduct a <u>real auction</u>. This art print will be sold to the highest bidder here today.*
>
> *Only one of these prints will be sold at this auction. After all bids have been collected, the person who is the highest bidder will be announced and he or she will be obligated to purchase the print at his or her bid price. We will accept cash or check for your purchase. We understand that you may not have anticipated the need to bring cash or your checkbook with you today, so we will also accept a signed Promissory Note payable on or before August 19, 1994. In any case, the highest bidder will be required to pay his or her bid amount and will then be able to take the art print home with him or her at the end of this session.*
>
> *Please understand you are participating in a <u>real auction.</u>*
>
> *Now take a few moments to determine the maximum dollar amount that you are prepared to pay for this art print. What is the most you are prepared to pay for this art print? I bid $______.*

IV. STATISTICAL PROCEDURES

The distributions of WTP we obtained are not normal. Absence of normality was confirmed using the Jarque-Bera normality test statistic, which is distributed as a chi-square with two degrees of freedom (Hall et al. 1990). Therefore, traditional two-sample t-tests are not appropriate and non-parametric or distribution free tests are required.

We employed three basic types of tests: (a) tests for significant differences in central tendency (mean or median WTP); (b) tests for statistical differences in the distribution of WTP; and (c) tests involving comparison of regression coefficients from the three sessions' WTP equations. The central tendency tests differ depending on whether independent or paired WTP estimates are being compared, as seen below.

Tests of Central Tendency for Independent Samples

The Mann-Whitney-Wilcoxon two-sample rank test is often used in the absence of normal distributions because the test only requires a continuous distribution (and the assumption of identical distribution shapes between groups). This non-parametric test determines whether the medians of two mutually independent random samples are significantly different. According to Gibbons (1993: 38-40), "The asymptotic relative efficiency of this test relative to the Student's t test is .955 for normal distributions, 1.00 for the continuous uniform distribution, and at least .864 for any continuous distribution..."

The Multi-Response Permutation Procedures (MRPP) are a series of non-parametric tests that make no distributional assumptions, not even asymptotically (Mielke 1984). The two independent samples being compared need not have the same shape or variance, or be symmetric. Permutation tests make efficient use of small sample sizes because exact p-values can be calculated. The underlying permutation distribution assigns equal probability to each possible permutation between observations in the two samples. The test statistic is equal to the sum of the weighted (among groups) average distance function values for all permutations in the samples.

If the sum of absolute median differences between the two samples is minimized, then a MRPP test is a test of the equality of medians. This test of medians is performed on the absolute magnitudes of the observations under analysis, rather than reducing the data to ordinal values as does the Mann-Whitney test of the medians. When the sum of the squared mean differences are minimized and mean differences are weighted by the degrees of freedom, MRPP is analogous to a two sample-t test of equality of means, but without the normality or asymptotic normality assumption. Given that the distributions associated with our data are non-normal, non-symmetric, and highly skewed, and that we have small sample sizes, we believe MRPP is an appropriate testing procedure for comparing independent open-ended WTP distributions.

Tests of Central Tendency for Paired Responses

In sessions 1 and 2, a hypothetical WTP question was followed by an actual cash WTP question. Although the paired WTP responses are not independent, they do provide a controlled comparison of the difference in WTP for a given respondent. Equality of the two paired responses was tested by several approaches. First, comparisons of WTP(h) and WTP(a_p) were performed using a paired sample permutation test similar to MRPP, from the set of Permutation Tests for Matched Pairs (PTMP). As with MRPP, the PTMP can be performed using the squared differences from the mean and absolute differences from the median. We believe PTMP is an appropriate test to use for paired samples for the same reasons we have chosen MRPP for mutually independent sample comparisons.

A second and more common test statistic we employed is the sign test of differences in the medians of paired data. This test does not make any assumption about the shape of the distribution other than it being symmetrical, but is less powerful than PTMP. In particular, the sign test has an asymptotic relative efficiency of at least .33 for any continuous symmetric distributions as compared to equivalent parametric tests (Gibbons 1993).

Finally, simple correlations between actual and hypothetical WTP were computed to examine the strength of linear relation between the paired vectors.

Tests of Differences Between WTP Distributions from Independent Samples

To test whether the distributions of open-ended WTP are different between hypothetical and actual WTP, we used a Kolmogorov-Smirnov non-parametric test. This is the same test statistic relied upon by Neill et al. (1994) in their comparison of hypothetical and actual WTP. The test involves a comparison of the distribution functions and calculation of the difference between the distributions. This test only requires that the distributions be continuous.

Tests of Regression Coefficients

The effectiveness of the reminder statement in discouraging respondents from reporting what they think the art print sells for was tested using a Wald test of the null hypotheses of $B_2=0$ in equation 1:

(1) $WTP(h{:}reminder) = B_o + B_1(MARKET) + B_2(SELL)$

where MARKET indicates how strongly respondents agreed or disagreed with the statement that they were in the market for art prints and SELL represents what respondents believed the market value of the print to be.

The hypothesis that *without* the reminder statement individuals' hypothetical WTP is influenced by what they think the print sells for is tested by ($B_2>0$) in equation 2, again using a Wald test:

(2) $WTP(h{:}no\ reminder) = B_o + B_1(MARKET) + B_2(SELL)$

The Wald test of the significance of the coefficients of equations 1 and 2 is recommended by Kennedy (1992, p. 61.) in the case of non-normally distributed residuals. The residuals of equations 1 and 2 are non-normally distributed. The Wald test is asymptotically distributed as a chi-square with degrees of freedom equal to the number of restrictions. If the SELL coefficient in the equation with no reminder is positive and significantly different from zero, but the SELL coefficient in the equation with the reminder is not, then the conclusion would be that what the respondents consider to be the fair market value of the print is a significant determinant of non-reminder hypothetical WTP, and that the reminder statement eliminates this response behavior.

The equality of the coefficients from the two hypothetical WTP equations 1 and 2, as well as from equation 3 regarding actual WTP:

(3) $WTP(a) = C_0 + C_1(MARKET) + C_2(SELL)$

was tested with likelihood ratio (LLR) tests of the equality of coefficients so as to provide a multivariate test of equality of the two types of valuation behavior. The null hypothesis of our LLR test was that the coefficients in the WTP equations for hypothetical and actual cash WTP were equal. The LLR test follows a chi-square distribution with degrees of freedom equal to the number of restrictions in the pooled regression minus the number of restrictions in the individual regressions.

IV. RESULTS

Before establishing that any differences between sessions are due to treatment differences (e.g., to the difference between real cash and hypothetical WTP), it was first determined that the respective samples were not statistically different in terms of standard demographics. To test for this across our sessions we performed one-way ANOVA's for education ($F=1.88$, $p=.16$), age ($F=1.4$, $p=.25$) and income ($F=.21$, $p=.81$) which showed that the samples are not statistically different at the .05 significance level. Sessions 1 and 3 consisted of about 75% percent women, but session 2 included only one male. We do not consider this problematic because multiple regressions indicated that gender was not a significant determinant of hypothetical or actual WTP. The high proportions of females reflects the nature of the population of university clerical and administrative staff that was sampled.

Table 1 reports mean and median WTP's for the three sessions. The independent means are $42, $26 and $14 for WTP(h:no reminder), WTP(h:reminder), and WTP(a), respectively. The paired estimates of actual WTP (those from sessions 1 and 2) are $12 and $13—very similar to the session 3 estimate.

Table 2 depicts the WTP distributions for the three treatments of the three sessions. Hypothetical WTP ranged up to $400 in session 1 and to $100 in session 2. Actual WTP ranged up to $40, $100, and $50 in sessions 1, 2, and 3, respectively. In session 2, the same person who bid $100 in the hypothetical treatment bid $100 in the actual treatment, and in fact paid this amount for the print. In session 1, only 7 of the 35 subjects bid the same amount in the hypothetical and actual treatments; and in session 2, 14 of the 33 subjects bid the same amount in both treatments. The general impression from Table 2 is that an actual WTP treatment garnered more bids from $1 to $10, and fewer bids above $40, than did the hypothetical WTP treatments.

Hypothetical versus Actual WTP

First, consider the tests across independent samples. As shown in the first row of Table 3, all four test statistics reject hypothesis 1, that WTP(h:no reminder) = WTP(a). Regarding hypothesis 2, that WTP(h:reminder) = WTP(a), the second row of Table 3 shows that the MRPP tests of the equality of means and equality of the medians both indicate equality must be rejected at the .05 significance level. However, the Kolmogorov-Smirnov test indicates that the distributions of WTP are not statistically different. The results of the Mann-Whitney test of the medians is in between, rejecting equality at .10 but failing to reject at .05. Thus, based on tests across independent samples, the evidence is mixed as to whether we reject the hypothesis that hypothetical and actual WTP are equal when the reminder statement is included.

Second, consider the tests of paired responses. As shown in the first row of Table 4, all three tests reject hypothesis 3, indicating that hypothetical WTP without a reminder was greater than actual WTP. And, as seen in the second row of Table 4, the three tests also reject hypothesis 4, indicating that hypothetical WTP with a reminder was also greater than actual WTP. In contrast to the equivocal results for hypothesis 2, the stronger tests of hypothesis 4 clearly suggest rejecting the claim that the reminder statement eliminated the disparity between hypothetical and actual WTP. Further tests presented in the Appendix also support rejecting the hypothesis of equal WTP estimates from session 2 respondents.

Effect of Reminder Statement

As seen in the third row of Table 3, all four tests indicate that we cannot reject equality of hypothetical WTP with and without the reminder statement. Although the reminder statement did lower the mean and median hypothetical WTP, we cannot conclude that the two measures of hypothetical WTP are different at the .05—or even the .10—significance level.

In terms of correlations, the paired responses from session 2 have a .733 correlation which is significant at the .001 level, whereas the responses from session 1 have a correlation of only .397, which is significant at the .05 level. Using a test described by Blalock (1972) these correlations are statistically different at the .05 level, indicating the reminder statement does significantly improve the relationship between actual and hypothetical WTP.

A test of the effectiveness of reminding respondents to report their WTP rather than what they think the art print sells for was performed by comparing the size and significance of the coefficient on the SELL variable in the WTP regressions of the three sessions. Non-normality in the residuals and dependent variables led us to test for coefficient significance with the Wald test. Since this restriction is linear, the F-statistics are valid (Kennedy 1992). F-statistics and their corresponding p-values are reported in equations 4, 5, and 6 for the hypothetical (without reminder), hypothetical (with reminder), and actual WTP treatments, respectively. Equation 4 is as follows:

(4)	WTP(h:no reminder) =	-43.91	+15.786(MARKET)	+.630(SELL)
	(F-statistic)	(3.87395)	(4.6699)	(23.6653)
	(F critical value)	(4.17)	(4.17)	(4.17)
	(P-value)	(.0578)	(.0383)	(.0000)

By the F and p-values, MARKET and SELL are both significant determinants of hypothetical WTP without the reminder statement, and the constant term is marginally significant as well. Most important, what the respondent thinks the print sells for has a significant influence on their reported WTP.

This contrasts with equation 5:

(5)	WTP(h:reminder) =	-5.578 +	8.886(MARKET)	+.085(SELL)
	(F-statistic)	(.31315)	(7.14514)	(3.22703)
	(F critical value)	(4.17)	(4.17)	(4.17)
	(P-value)	(.5802)	(.0124)	(.0832)

By the F and p-values of equation 5, MARKET is a significant determinant, but SELL is not a significant determinant, of hypothetical WTP with the reminder statement. This can now be compared with equation 6 for actual WTP:

(6)	WTP(a) =	-.096 +	4.363(MARKET)	+.046(SELL)
	(F-statistic)	(.00032)	(6.32786)	(1.59605)
	(F critical value)	(4.17)	(4.17)	(4.17)
	(P-value)	(.9859)	(.0177)	(.2165)

By the F and p-values of equation 6, MARKET is a significant determinant, but SELL is not a significant determinant, of actual WTP. Thus, equations 5 and 6 have the same pattern of significance on the regression coefficients: the constants are not significantly different from zero, MARKET is a significant determinant of WTP, and SELL is not a significant determinant of WTP. The reminder statement does appear to aid respondents in behaving more like they do in the actual cash market, where what they think it sells for has no statistically significant effect on WTP.

LLR tests of the equality of the coefficients of equations 4 and 6, 5 and 6, and 4 and 5 yield chi-squares of 18.686, 76.50, and 41.89, respectively. The critical chi-square value for each of these tests is 7.82, indicating that the independent variables do not affect the dependent variables in the same way across equations. Once again, actual cash behavior and hypothetical WTP behavior are different. Though inclusion of the reminder statement does make the hypothetical WTP more closely mimic the actual cash WTP (as indicated by the similarity of coefficients on the SELL variable), the overall valuation behavior exhibited in the hypothetical WTP equations is different from the behavior exhibited by the actual cash WTP equation.

Separate versus Paired Actual WTP

As seen in the last two rows of Table 3, hypotheses 6 and 7 are not rejected. This is, of course, not surprising given the similarity of the three actual WTP responses (Table 1). We conclude that respondents' actual WTP responses were not influenced by their prior hypothetical WTP responses.

DISCUSSION

Our results are consistent with those of previous experiments comparing hypothetical and actual WTP for a market good, in that hypothetical WTP overestimated actual payment. Depending on the wording of the hypothetical treatment, we found mean hypothetical WTP to be 2 or 3 times mean actual WTP for the art print. Even our rather labored reminder to the hypothetical WTP respondents failed, based on most statistical tests, to avoid over-estimating actual WTP.

Use of the reminder statements lowered hypothetical WTP, but the reduction was not statistically significant. However, the reminder did result in a significant improvement in the correlation of actual to hypothetical WTP. Further, the regression analyses showed that the reminder lessened the association of hypothetical WTP with perceived market price, apparently causing the effect of perceived market price on hypothetical WTP to be closer to its effect on actual WTP. The principal differences between the hypothetical bids received with versus without the reminder statement are that with the statement there were fewer very high bids and more $0 bids (Table 2). In fact, the two highest bids obtained in the without-reminder treatment ($150 and $400) account for much of the difference in mean WTP between the two treatments. Thus, an advantage of the reminder may be that it reduces the tendency of some respondents to give unrealistically large WTP bids.

We found that perceived market price of the print was significantly related to hypothetical WTP without the reminder, but essentially unrelated to actual WTP. Some subjects apparently relied on their estimates of market price to help estimate their WTP in the hypothetical treatment. This finding raises the possibility that experiments using market goods do not provide a reasonable test of criterion validity of CV on nonmarket goods. It is feasible that hypothetical WTP will more closely approximate actual WTP where price cues do not affect hypothetical WTP. Of course, *other* unintended cues may affect the chances of CV to estimate WTP for nonmarket goods—but that possibility does not necessarily excuse the use of market good experiments to test the validity of CV.

The three estimates of mean actual WTP obtained from our three subject groups were very similar—about $13. The hypothetical treatments had essentially no effect on the subsequent actual cash bids. Most subjects' actual cash bids were considerably lower than their prior hypothetical bids. Subjects apparently were not bothered by the discrepancy between their two bids, as none objected to being asked for actual WTP after they had already provided hypothetical WTP.

Our wording in the hypothetical treatment without reminder was very similar the open-ended versions used by Neill et al. (1994). And our mean actual payment (about $13) was similar to Neill et al.'s (about $9 for the painting and $12 for the map). The ratio of hypothetical to actual payment obtained by Neill et al. was 4:1 for the painting, but 9:1 for the map (and that is excluding WTP's over $1,000), compared with our 3:1 with no responses excluded. One obvious difference between our methods and those of Neill et al. is that our subjects were university staff instead of students. Unfortunately, we have no way to know for sure whether this or some other difference accounted for the quite different results between the two studies.

CONCLUSIONS

Both independent sample and paired sample comparisons indicate that hypothetical WTP exceeds actual cash WTP for the art print. The use of reminder statements reduced the difference between hypothetical and actual WTP from 3:1 to 2:1. Although the reduction was not statistically significant, the reminder did result in a significant improvement in the correlation of hypothetical to actual WTP, and in a reduction of the association of hypothetical WTP to perceived market price to approximate the association found between actual WTP and perceived price.

In addition to confirming previous findings that hypothetical WTP tends to over-estimate actual WTP, our findings support two other conclusions. First, validity experiments using market goods may not be directly relevant to estimating WTP for nommarket goods because of the effect of price cues. Second, prior hypothetical WTP estimates do not appear to affect mean actual WTP from the same subjects. Thus, we have no evidence to find fault with other studies that have used paired WTP responses to evaluate the veracity of hypothetical WTP.

What can be learned about criterion validity experiments? There are at least two avenues to pursue. One is to continue to debrief and probe respondents about the differences in their decision processes in hypothetical versus actual cash decisions. Using this knowledge, statements to combat the hypothetical nature and place them in an actual payment frame of mind can be developed to improve the match between intended behavior and actual behavior. Our study made some progress in this regard, and more refinement may be possible. Second, CV researchers may wish to adopt the viewpoint of market researchers who face a similar dilemma with intended purchase behavior: calibration. That is, we can begin to assess how much of the hypothetical WTP is "noise" and how much is actual cash "signal". NOAA and DOI in recent proposed CV rules (NOAA 1994; DOI 1994) suggests a calibration factor of .5, but requests empirical evidence as to whether .5 or any other number is valid. Our study suggests some calibration factor may be necessary, but the magnitude appears to depend on the details of the CV survey such as question wording and question format. Clearly, numerous replications with different goods, different question formats and larger samples are warranted before we have a sense as to the range of calibration factors that might be credible. The long term goal of such research would be to develop bias functions that would assist us in estimating the ratio of signal to noise for different types of natural resources and survey designs.

REFERENCES

Bishop, Richard and Thomas Heberlein. 1979. Measuring Values of Extra-Market Goods: Are Indirect Measures of Value Biased? American Journal of Agricultural Economics 61(5):926-930.

Blalock, H. 1972. *Social Statistics* (2nd Edition). McGrawHill, New York.

Boyce, R. R., T. C. Brown, G. D. McClelland, G. L. Peterson, and W. D. Schulze. 1989. "Experimental evidence of existence value in payment and compensation contexts." In *Western Regional Research Project W-133, Second interim report, Benefits and costs in natural resource planning*, Department of Agricultural and Resource Economics, University of Maine, Orono, K. J. Boyle and T. Heekin, editors, pp. 305-336.

Brookshire, D., and D. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review* 77:554-566.

Carson, R.T., J.L Wright, N.J. Carson, A. Alberini, N.E. Flores. 1994. A Bibliography of Contingent Valuation Studies and Papers. Natural Resource Damage Assessment Inc., La Jolla, CA.

Champ, P., R. Bishop, T. C. Brown and D. McCollum. 1994. "Some Evidence Concerning the Validity of Contingent Valuation: Preliminary Results of an Experiment." In R.Ready, compiler, *Benefits and Costs Transfers in Natural Resources Planning*, Department of Agricultural Economics, University of Kentucky, Lexington.

Coursey, D., W.Schulze and J.Hovis. 1987. "The Disparity Between Willingness to Pay Measures of Value." *Quarterly Journal of Economics* 102:679-690.

Cummings, R., G. Harrison and E.E. Rutstrom. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic*

Review 85 (in press).

Department of Interior. 1994. "Natural Resource Damage Assessments: Proposed Rule." *Federal Register* 43CFR, 59(85): 23098-23111. May 4, 1994.

Duffield, J., and D. Patterson. 1992. "Field Testing Existence Values: Comparison of Hypothetical and Cash Transaction Values." In Bruce Rettig, compiler, *Benefits and Costs in Natural Resource Planning*, Fifth Interim Report, Department of Agricultural and Resource Economics, Oregon State University, Corvallis, OR.

Gibbons, J. D. 1993. *Nonparametric Statistics, an Introduction.* Series: Quantitative Applications in the Social Sciences. Sage Publications. Thousand Oaks, CA.

Hall, Robert, Jack Johnston and David Lilien. 1990. Micro-TSP. QMS, Irvine, CA.

Kealy, M. J., and R. Turner. 1993. "A Test of the Equality of Close-Ended and Open-Ended Contingent Valuations." *American Journal of Agricultural Economics* 75(2): 321-331.

Kealy, M. J., J. Dovidio and M. Rockel. 1988. "Accuracy in Valuation is a Matter of Degree." *Land Economics* 64(2): 158-171.

Kennedy, P. 1992. *A Guide to Econometrics* (3rd Ed.) The MIT Press, Cambridge, MA.

Loomis, J. 1989. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." *American Journal of Agricultural Economics* 71(1):76-84.

Mielke, P., Jr. 1984. "Meteorological Applications of Permutation Techniques based on Distance Functions." Pages 813-830 in *Handbook of Statistics*, Vol 4, in P.R. Krishnaiah and P.K. Sen, eds. Elsevier Science Publishers.

Mielke, P., and K. Barry. 1994. "Permutation Tests for Common Locations Among Samples with Unequal Variances." *Journal of Educational and Behavioral Statistics*, 19(3):217-236.

Neill, H., R. Cummings, P. Ganderton, G. Harrison, and T. McGuckin. 1994. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(2):145-154.

National Oceanic and Atmospheric Administration. 1994. "Natural Resource Damages: Proposed Rules." *Federal Register* 15CFR, Part 990. Vol 59(5):1062-1191. January 7, 1994.

Seip, K. and J. Strand. 1992. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payments." *Environmental and Resource Economics* 2:91-106.

Welsh, Michael. 1986. Exploring the Accuracy of the Contingent Valuation Method: Comparison with Simulated Markets. Unpublished Ph.D. dissertation, University of Wisconsin.

APPENDIX

If the WTP(a_p) that individual i gives in the follow-up response is equal to his or her original WTP(h), then $B_0=0$ and $B_1=1$ in equation A1 which is for the first session, and in equation A2, which is for the second session:

(A1) $\quad WTP(a_p)_i = B_0 + B_1[WTP(h\text{:no reminder})_i]$

(A2) $\quad WTP(a_p)_i = B_0 + B_1[WTP(h\text{:reminder})_i]$

This test is equivalent to a paired t-test but the approach also gives useful estimates of coefficients. Loomis (1989) used this test for comparison of a respondent's original answer and retest answer eight months later. Typically, the significance of the coefficients would be tested using a t-test, but non-normality of our data suggests use of a more general test such as the Wald test.

Regression for equation A2 yielded:

(A3) $\quad WTP(a_p) = -1.385 + .5634(WTP(h\text{:reminder}))$

The Wald test of $B_0=0$ yields a p-value of .6872, indicating that the constant is not significantly different from zero. The Wald test of $B_1 = 1$ yields a p-value $< .0001$, indicating that the coefficient on WTP(h) is significantly different from one. We also tested the null hypothesis $B_1=0$; the Wald test yields a p-value $< .0001$, indicating that the slope coefficient is significantly different from zero. We interpret this to mean that although B_1 does not equal 1 with the reminder statement, the coefficient is still significant and every dollar of WTP(h) translates at the margin into \$.56 of actual WTP.

The regression for equation A1 yielded:

(A4) $\quad WTP(a_p) = 8.79 + .067(WTP(h\text{:no reminder}))$

The Wald test of $B_0=0$ yields a p-value $< .0001$, indicating that the constant is significantly different from zero and a p-value $< .0001$ for $B_1=1$, indicating that the slope coefficient is significantly different from one. The Wald test of $B_1=0$ yields a p-value of .0130, indicating that the slope coefficient is significantly different from zero. Again, although B_1 does not equal 1 without the reminder statement, the coefficient is still significant: every dollar of hypothetical WTP translates at the margin to \$.07 of actual WTP.

The reminder statement appears to be effective in narrowing the gap between hypothetical and actual WTP responses, as indicated by the increase in B_1 (from .07 to .56) and insignificance of B_0 (e.g., compare the coefficient on B_1 when the statement is included. However, these results suggest rejection of $B_1 = 1$ for both equations, indicating that hypothetical and actual WTP were not equal for paired responses, whether a reminder was included or not.

TABLE 1
Comparisons of Hypothetical and Actual WTP

Session	Sample size	Mean (median) WTP by Treatment [S.E. of mean]			Ratio: WTP(h) /WTP(a)
		1. Hypothetical: no reminder	2. Hypothetical: with reminder	3. Actual	
1	35	42.34 (25) [11.38]		11.63 (6) [1.92]	3.64
2	33		26.29 (20) [4.45]	13.42 (6) [3.42]	1.96
3	32			14.48 (10) [2.27]	

TABLE 2
Number of Bids Received by WTP Class

WTP class ($)	Treatment				
	Session 1		Session 2		Session 3
	WTP (h:no reminder)	WTP(a)	WTP(h: reminder)	WTP(a)	WTP(a)
0	2	9	6	7	3
1-10	6	11	7	15	15
11-20	7	8	6	3	6
21-30	5	5	4	4	5
31-40	4	2	3	2	1
41-50	6	0	2	1	2
51-99	3	0	4	0	0
100	0	0	1	1	0
>100	2[1]	0	0	0	0

[1] These bids were for $150 and $400.

TABLE 3
Probability Levels from Tests across Independent Samples

Hypothesis	Statistical test			
	Mann-Whitney (test of medians)	MRPP (test of means)	MRPP (test of medians)	Kolmogorov-Smirnov (test of distributions)
1. WTP(h:no reminder) = WTP(a)	.0017	.0015	.0037	.009
2. WTP(h:reminder) = WTP(a)	.0917	.0216	.0327	.318
5. WTP(reminder) = WTP(h:no reminder)	.2071	.218	.314	.511
6. WTP(a) = WTP(a_{p1})	.2945	.785	.390	.569
7. WTP(a) = WTP(a_{p2})	.2298	.355	.376	.475

TABLE 4
Probability Levels from Tests across Paired Responses

Hypothesis	Statistical test		
	Sign test (of medians)	PTMP (test of means)	PTMP (test of medians)
3. WTP(h:noreminder) = WTP(a_p)	.0000	.0003	.0000
4. WTP(reminder) = WTP(a_p)	.0007	.0013	.0004

A Comparison of Contingent Values and Actual Willingness to Pay using a Donation Provision Mechanism with Possible Implications for Calibration

Patricia A. Champ
Richard C. Bishop
Thomas C. Brown
Daniel W. McCollum[1]

May 12, 1995

[1] Champ, Brown and McCollum are with the Rocky Mountain Forest and Range Experiment Station, USDA Forest Service, Fort Collins, CO. Bishop is with the Department of Agricultural Economics, University of Wisconsin.

A Comparison of Contingent Values and Actual Willingness to Pay using a Donation Provision Mechanism with Possible Implications for Calibration

This paper expands on an earlier investigation of the validity of contingent values using a donation provision mechanism (Champ et al. 1994). Reanalyzing our earlier data and incorporating new data, we explore the overestimation by contingent values of actual willingness to pay (WTP).

The Provision Mechanism

The possibility of free riding means that donation mechanisms are less than ideal provision mechanisms for contingent valuation (CV) studies. In more technical terms, such mechanisms are not "incentive compatible." Still, it is difficult to know how seriously to take this disadvantage. Whether individuals perceive the incentives to reveal (or not reveal) their true WTP, and in turn respond to those incentives is not clear. The research on free riding, for example, suggests that it occurs less frequently and with less effect than economic theory predicts (Marwell and Ames 1981, Schneider and Pommerehne 1981, Brubaker 1982, and Christianson 1982). Furthermore, donation mechanisms have certain advantages that should be considered before rejecting them out of hand. Bishop and McCollum (1995) point out that the content validity of CV studies is enhanced if the scenario is both acceptable and believable to respondents. Because donations are voluntary, scenario acceptance may be more easily won than when tax or other unpopular payment vehicles are used. Furthermore, for some types of interventions, donation mechanisms may be more plausible than other types of provision mechanisms. For example, respondents may find referenda for small projects implausible. That they might be asked to donate money may be more believable in such cases. Finally, more applied research has been conducted to investigate the validity of donation mechanisms for providing public goods. Examples of such experiments include Duffield and Patterson (1992) and Seip and Strand (1992). To date, no comparable criterion validity work had been done for referenda. To the extent that calibration of CV results becomes necessary, laboratory and field experiments that measure both actual cash transactions and contingent values for the same environmental amenity are among the more promising approaches.

The Good

The goal of this study was to learn more about the relationship between contingent values and actual WTP and we chose an environmental good suitable for such a comparison. Most important, we wanted the amenity to have a potentially large share of its total value be nonuse value, as this is the area where the validity of CV is currently being most intensively questioned (Cambridge Economics, Inc. 1992). We also needed a good that could be reasonably purchased by respondents, thus generating value estimates based on actual cash transactions. Finally, we wanted the good to be divisible so there would be a relationship between the amount each individual paid and provision of the good. We chose a program run by the National Park Service at Grand Canyon National Park to remove 40 miles of compacted dirt roads on the North Rim of the Grand Canyon. The North Rim is not open all year and requires driving 215 miles by road from the South Rim. As a result, few visitors to the Grand Canyon actually visit the North Rim. The National Park Service would like to remove the roads and ultimately designate the area where the roads are located a Wilderness Area. Volunteers are available to provide the labor but the Grand Canyon National Park did not have funding to pay for food and supplies for those volunteers. It costs approximately $640 to remove one mile of road, based on the price of food and supplies for the volunteers. At the time of this study, our project was the only source of funding for the road removal program.

The Data

The study treatments are summarized in Table 1. Data were collected over a period of two years. Separate samples of Wisconsin residents drawn from the same sample frame participated in mail surveys in October of 1993 and 1994. In 1993, 1700 surveys were mailed. Eight hundred fifty of those surveys posed a dichotomous-choice (DC) question which gave respondents the opportunity to actually donate a specified amount for road removal.[2] This treatment will be designated AC (for "actual cash") in what follows. The other 850 were asked a parallel CV question about whether or not they would be willing to donate a specified amount if given the opportunity to do so. In 1994, the CV data set was extended by duplicating two offer amounts used in 1993 ($15 and $50) and adding offer amounts of $75, $100, $150, and $200.[3] The combined 1993 and 1994 data sets will be labeled DC (for simple "dichotomous-choice"). In 1994, two treatments were added. One treatment, which we refer to as polychotomous-choice (PC), posed the same CV question as the DC treatment, but the response categories were expanded from two (yes and no) to six (definitely yes, most likely yes, not sure but leaning toward

[2]The WTP question asked individuals to pay a specified amount but many individuals decided to send a check for an amount other than the specified amount. In particular, several people who were asked to pay $1 sent a check for $10.

[3]The two duplicate offer amounts were used to test whether the 1993 and 1994 data were comparable. We found that the distributions of yes and no responses to each of the offer amounts were not significantly different between the two years.

yes, not sure but leaning toward no, most likely no, and definitely no). Ready, Whitehead, and Blomquist (1995) were the first to use the PC response format with a closed-ended CV question. However, they labeled the response categories differently than we did in this study. The other treatment posed a DC CV question and followed up with a question asking respondents to rate, on a ten-point scale, how certain they were that they would actually send a check (or not send a check if they said no to the DC CV question) if asked to do so. Only the endpoints of the scale were labeled, with one corresponding to "very uncertain" and ten corresponding to "very certain." For simplicity, we will refer to this treatment at the DCWC ("dichotomous choice with certainty" question). The WTP questions, the rest of the questionnaires, and all other survey materials such as cover letters were designed to be as similar as possible for all treatments. The questionnaires included questions about demographic background, experience with National Parks, and attitudes toward the environment in general and wilderness in particular.

Table 2 shows the offer amounts, the initial sample sizes, the number of returned surveys, and response rates for each treatment. Given that the treatment groups were randomly selected from the same population, one would expect the groups to respond in a similar manner to objective or factual questions. However, there was a significant difference in response rate between some of the treatments (i.e., 51% for the CV treatment in 1993 and 44% for the actual payment group in 1993). Contingency table analysis was used to test to investigate response rate effects among the various treatments. Results of that analysis suggested the various treatment groups are representative of the same population and any differences among the WTP of the various treatments can be attributed to factors other than sampling or response rate effects.[4]

[4] See Champ (1994) for detailed description of analyses of differences between the two 1993 treatments.

Therefore, the data from the simple DC treatments in 1993 and 1994 were combined for the following analyses.

Actual versus Hypothetical Payments

We first compare the AC and DC treatments. At all offer amounts where we had both CV and actual payment results, respondents to the actual payment question were less likely to respond positively to the WTP question than CV respondents. Likewise, the estimated mean WTP from the simulated market data is significantly less than the mean WTP from the CV data set (Table 3).

One means of providing insight into the overestimation of actual WTP by contingent values is to differentiate respondents who would actually pay in a manner consistent with their CV response if they had been asked to do so from those who would not. The following procedures were developed to distinguish DC respondents who are inconsistent in the sense of responding differently to the CV question than we predict they would if asked to actually pay. First, using the AC data, a WTP function was estimated. Table 4 describes the variables and the estimated model.[5] The model fits the data for the AC group quite well; 86% of the responses are predicted correctly. Most important, this model predicts 81% of the yes responses correctly. Estimated coefficients from the AC-based model were then applied to the DC data to predict the probability that an individual who answered all the requisite questions would respond positively if asked to

Admittedly, this model has many explanatory variables but the purpose of the model is to predict actual payment responses as accurately as possible. Given this goal, the model is appropriate. The collinearity among the explanatory variables may in part be responsible for the predictive strength of the model. However, this collinearity may result in estimated coefficients that do not reveal the nature of the individual relationship between a specific explanatory variable and the response to the WTP question. Those individual relationships are analyzed extensively in Champ (1994) using contingency table analysis.

actually pay the amount posited in the WTP question.[6] DC respondents with a predicted probability greater than 0.5 were predicted to say yes to an actual payment WTP question and all others were predicted to say no.

It is encouraging evidence of the predictive strength of the model that the overall percentage of the AC respondents saying yes to the actual WTP question is similar to the percentage of DC respondents that were predicted to say yes based on the model. DC respondents who said yes to the CV question but which our model predicted would say no if actually asked to pay were classified as "inconsistent." Likewise individuals who said no to the CV question and were predicted as saying yes to the actual payment question were also classified as inconsistent. Individuals whose answers to the CV question conformed to model predictions were designated as "consistent." See Table 5 for the predicted response to actual payment versus the observed response to the CV question.

The results of the consistency analysis are enlightening with respect to the source of overestimation of actual WTP by contingent values. Twenty-six percent of the respondents to the CV question were designated to be inconsistent. Only nine respondents were inconsistent in the sense of saying no to the CV question when the model predicted they would say yes if payment were real. The rest of the respondents who were inconsistent said yes to the CV question when the model predicted they would not say yes if payment were real (Table 5).

Comparison of consistent and inconsistent respondents provides some insight into why contingent values overestimated actual WTP in our case. The offer amount appears to affect

[6]In other words, only cases which did not have missing data on any of the independent variables in the model could be used. There were 458 useable cases out of the initial DC sample of 648.

whether an individual saying yes to the CV question was inconsistent with the model prediction. Larger percentages of the yes-respondents were inconsistent as the offer amount increased (Table 6). However, as a percentage of both the yes and no responses, the percentage of respondents that were inconsistent was fairly constant across the offer amounts. Unfortunately, the measures collected in this study did not allow us to investigate the possible causes of this observed phenomena.

Because most of the people who were inconsistent said yes to the CV question, the following analysis focuses on them.[7] Comparing the responses of individuals, who were consistent to those who were not shows several differences between the two groups. Responses to statements about why individuals answered yes to the CV question show more consistent than inconsistent respondents circled "definitely true" to the statement that the road removal program might be worth the amount they were asked to pay. Likewise more consistent respondents said it was "definitely" or "somewhat" true that the total number of feet of road that would be removed was important in their decision to pay. Consistent respondents were more likely to respond "definitely false" to the statement "I would rather see the money go to a better project." Consistent respondents were also more likely to say it is "definitely true" that it is important that the area be designated 'wilderness' after the roads are removed. Furthermore, consistent respondents were more likely to have visited a National Park in the past and think it is very likely they will visit the North Rim of the Grand Canyon in the future. Consistent and inconsistent respondents also have some different demographic characteristics, with inconsistent ones being

[7]If we compare inconsistent response to consistent response for both yes and no responses, the comparison is very similar to looking at who said yes and who said no to the DC CV question.

more likely female, living in rural areas and having lower education levels and incomes. It is also interesting to note the ways in which consistent and inconsistent respondents who said yes to the CV question were similar. Both groups had similar attitudes toward the environment and toward wilderness areas in particular. These similarities suggest that differentiating between consistent and inconsistent yes responses to the CV question is rather subtle. Therefore, an approach to calibrating which calls for the researcher to differentiate between consistent and inconsistent respondents without additional information will be quite challenging. We suggest that it may be more effective to have respondents to the CV question identify their responses as consistent or inconsistent (although not necessarily in those words) and the researcher can use this information to calibrate.

Returning to a broader perspective, it is worth emphasizing that 74 percent of the DC respondents were consistent. Critics of CV often argue that people in general cannot respond to CV questions in a meaningful way. The preceding analysis suggests that most individuals who say no to a hypothetical WTP question would really not pay if asked to do so. While some of the people who say yes to the CV question would respond the same way if payment were real, others would not.

Alternative Response Formats

The additional treatments in 1994 were needed to explore possible methods for correcting the overestimation of actual WTP that occurred in the DC treatment. Based on the results of the consistency analysis, we hypothesized that, at least for donation mechanisms, many people who say yes to a simple DC CV, but are uncertain about whether they would actually pay, would not actually contribute the money. The PC and DCWC treatments were designed to allow respondents to express uncertainty. We implicitly assumed that individuals who are uncertain about how to respond to a closed-ended CV question know that they are uncertain and will reveal their uncertainty to the investigator if allowed to do so.

To investigate whether data from the PC and DCWC treatments could be used to reduce the tendency toward overestimation, we followed two approaches. First, we coded people who revealed uncertainty about their yes answers, either by answering "most likely yes" or "not sure but leaning toward yes" in the PC treatment or by circling any value less than 10 ("very certain") on the follow-up certainty question in the DCWC treatment as if they answered no. For convenience those who fell into these categories will be designated as "uncertain respondents," but it should be emphasized that we were only interested in respondents who answered the CV question in a positive way and revealed some uncertainty about that response through either the PC response format or the follow-up certainty question. This then allowed us to compare the distributions of values for PC and DCWC treatments where uncertain yes responses were interpreted as no with the distribution of values for the AC treatment. Second, we tested to see whether those who revealed uncertainty in the ways just indicated would be comparable to those who were earlier identified as inconsistent in answering yes in the DC treatment.

Comparing the percentages of respondents in each treatment who said yes to the WTP question (Table 3) suggests that the percentage saying yes are not significantly different for the AC, PC, and DCWC treatments at the offer amounts of $15 and $50. However, the percentages yes for all three of these treatments are significantly less than the percentage of respondents saying yes in the DC treatment. Table 3 also reveals that the mean WTP estimated based on the data from the AC, PC, and DCWC data are not significantly different. The estimated logistic functions based on various ways of dichotomizng the DCWC data are graphed in Figure 1. The AC function is also included in Figure 1 as a benchmark. Depending on how one chooses to interpret the responses to the follow-up certainty question, functions which vary from a lower bound of close to the function based the AC data to an upper bound based on the uncalibrated responses to the DC CV question. A similar pattern is shown in Figure 2 with the PC data. However the dichotomization which calls for coding only definitely yes responses as positive responses to the WTP question, does not show a function similar to the benchmark function based on the AC data. The quality of the PC data is questionable given the percentage of yes responses increases significantly from \$15 to \$50 and the resulting large confidence interval around the estimate of mean WTP.

Next consider the relationships between those who expressed some uncertainty but answered positively in the PC treatment respondents in the DC treatment that were classified as inconsistent because they responded yes to the CV question but we predict they would not actually pay if asked to do so. Comparisons were limited to the \$15, \$50 and \$75 cells because those are the only cells where the DC and PC data sets overlap. On most measures elicited in this survey, the responses of the two groups are distributed similarly. However, there are a few significant differences that suggest these two groups may not be representative of the same population. First, the distribution of responses to two of the statements following the CV question were significantly different, with more inconsistents than uncertain PC respondents circling definitely false to the statements "My decision about whether or not to pay was based on the number of feet of road that would be removed if I agreed to pay" and "When I was deciding whether to pay, I considered the fact that I already pay for environmental projects through taxes."

Another significant difference (at the 10% level) was that more inconsistent respondents than uncertain PC respondents said they had visited the Grand Canyon National Park in the past.

These differences seem slight but they did not occur when comparing inconsistent members of DC with uncertain DCWC respondents. Again, such comparisons were only possible for those assigned to the $15, $50 and $75 cells. The distributions of responses to nearly all the questions in the questionnaire are similar for the two groups. The only variable for which the responses of those two groups were distributed significantly different was whether they had visited the Great Smokey Mountains National Park, a result we interpreted as a statistical fluke. All other variables were determined to have similar distributions for those two treatment groups based on the results of contingency table analysis.

Given the results of this study, we suggest that the follow-up certainty question seems to have potential as a means of calibrating DC CV responses when a donation mechanism is used. The cost of including the follow-up certainty question in a CV survey is relatively low. The DC CV data is still available if the researcher decides to not use the information in the follow-up question and the results from this study suggest that the follow-up question did not significantly affect the response rate or distribution of responses to the DC CV question relative to the DC treatment without the follow-up question. The same cannot be said for the PC data. The analyses using the PC data suggest that this format does not provide a lower bound that is distributed like the AC data. Furthermore, the upper bound distribution provided by the PC data appears to be different from that of the DC CV data.

Conclusion, With Emphasis on Calibration

As we emphasized at the outset, use of a donation mechanism to estimate WTP raises serious theoretical questions that should not be ignored. The much lower value from the AC treatment lends itself easily to a free-rider interpretation. Many people who did not send us a check for Grand Canyon road removal may simply have been free riders, hoping that other would pay. That the DC treatment got a higher value might indicate that the incentives to free ride were less potent where respondents do not actually have to write checks. Thus, our conclusions about the validity and usefulness of donation mechanisms could be quite negative. If so, then using the AC treatment as a basis for calibration would be a doubtful proposition.

Such an interpretation is overly critical in our opinion. Consider a different argument. If some respondents really send in checks despite the free-ride incentive and if no other perverse incentives in the opposite direct are identified in the experimental design, then values based on responses to the AC treatment lend themselves well to interpretation as lower bounds on the values that would be arrived at, all else equal, if an incentive compatible mechanism could have been used. However, incentive compatibility may come at a price in terms of content validity. As we pointed out at the outset, a referendum in this particular case runs risks of scenario rejection due to the need to use a less acceptable payment vehicle (e.g., taxes) and implausibility for such a small project.[8]

Suppose one is trying to investigate the benefits of the some small project like Grand Canyon road removal and that, as will often be the case, actual cash transactions are not feasible. Suppose further that later studies replicate the kinds of results for donation mechanisms that we have reported here. Such an investigator would not face a simple choice. One option would be

---

[8]Specifically, respondents may not believe that an actual referendum would be conducted to decide whether to fund a project that costs $25,000 in a National Park.

to use a donation vehicle, but attempt to calibrate the responses through some mechanism like those used in our PC or DCWC treatments. Such a study would produce a lower bound estimate of benefits with a substantial amount of validity based on past studies that had compared values based on actual donations with values estimated using the CV donation mechanism. The other option would be to employ an incentive compatible CV mechanism such as a referendum. The referendum would produce results with a theoretical edge in the area of incentive compatibility, but that would remain, at least given our current state of knowledge, uncalibrated. At least at present, such referendum values could not fall back on studies involving comparisons with actual cash values to support their validity as either estimates of the true value or as lower bounds on true values.

On this basis, we would conclude that donation mechanisms remain a useful tool for CV studies. Definitive conclusions about validity and calibration are not going to come from any one study. Nevertheless, the work summarized here is suggestive of a line of future research that may prove fruitful in estimating lower bounds on values of environmental amenities with substantial validity.

REFERENCES

Bishop, R.C. and D.W. McCollum. Forthcoming. "Assessing the Content Validity of Contingent Valuation Studies." Proceedings from the 1995 Western Committee on the Benefits and Costs of Natural Resource Planning (W133) Meeting, Monterey, CA.

Brubaker, E. 1982. "Free Ride, Free Revelation or Golden Rule?" *Journal of Law and Economics* 18(1):147-159.

Cambridge Economics, Inc. 1992. A symposium titled "Contingent Valuation: A Critical Assessment." Washington, D.C.

Champ, P.A. 1994. Nonmarket Valuation of Natural Resource Amenities: A Validity Test of the Contingent Valuation Method. Ph.D. dissertation, Department of Agricultural Economics, University of Wisconsin, Madison, WI.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1994 "Some Evidence Concerning the Validity of Contingent Valuation: Preliminary Result of an Experiment." In R.C. Ready, W-133 Seventh Interim Report. Western Regional Research Publication, Lexington, KY.

Christiansen, G.B. 1982. "Evidence for Determining the Optimal Mechanism for Providing Collective Goods." *American Economist* 26(1):57-61.

Duffield, John W., and David A. Patterson. 1992. "Field Testing Existence Values: Comparison of Hypothetical and Cash Transaction Values." Paper presented at the joint meetings of the Western Committee on the Benefits and Costs of Natural Resource Planning (W133) and the Western Regional Science Association, South Lake Tahoe, Nevada.

Marwell, G., and R.E. Ames. 1981. "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods." *Journal of Public Economics* 15:295-310.

Ready, R.C., J.C. Whitehead, and G.C. Blomquist. Forthcoming. "Contingent Valuation when Respondents are Ambivalent." *Journal of Environmental Economics and Management.*

Schneider, F. And W.W. Pommerehne. 198. "Free Riding and Collective Action: An Experiment in Public Microeconomics." *Quarterly Journal of Economics* 97:689-702.

Seip, K. And J. Strand. 1992 "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment." *Environmental and Resource Economics* 2(1):91-106.

Table 1: Summary of Study Treatments

Year	Symbol	Nature of the Treatment
1993	AC	Actual cash donated; dichotomous choice vehicle; no follow-up certainty question.
1993, 1994	DC	CV; dichotomous choice; no follow-up certainty question.
1994	PC	CV; polychotomous choice; no follow-up certainty question.
1994	DCWC	CV; dichotomous choice with follow-up certainty question.

Table 2: Offer amounts, sub-sample sizes, responses, and response rates

Offer Amount	AC	DC (1993)	DC (1994)	PC	DCWC
	Number of Surveys Mailed / Number of Useable Surveys Returned				
$1	125 / 50	125 / 56	*	*	*
$5	175 / 66	175 / 82	*	*	*
$8	175 / 68	175 / 75	*	*	*
$12	125 / 60	125 / 65	*	*	*
$15	125 / 38	125 / 59	100 / 45	100 / 48	100 / 35
$50	125 / 50	125 / 56	100 / 42	100 / 34	100 / 42
$75	*	*	100 / 35	100 / 39	100 / 35
$100	*	*	100 / 43	*	*
$150	*	*	100 / 50	*	*
$200	*	*	100 / 40	*	*
Response Rate[4]	44%	51%	47%	47%	42%

[1]DC refers to dichotomous choice format.
[2]PC refers to polychotomous choice format.
[3]DCWC refers to the dichotomous choice question with the follow-up question about how certain the respondent felt about her response to the CV question.
[4]Response Rate = ((Number Complete)/(Number Mailed - Number Undeliverable)) x 100.
*Data not collected.

Table 3 : Percentage Yes Responses to WTP Question by Treatment and Offer Amount

	AC	DC	PC[1]	DCWC[2]
$1	24%	53%	*	*
$5	15%	51%	*	*
$8	25%	39%	*	*
$15	13%	46%	13%	23%
$50	4%	27%	3%	12%
$75	*	31%	13%	3%
$100	*	19%	*	*
$150	*	18%	*	*
$200	*	28%	*	*
Estimated WTP Based on Available Data	$9.18	$78.79	$15.93	$13.95
95% Conf. Interval	[6.08,24.38]	[58.34,128.14]	[2.27,345.98]	[8.44,28.72]

[1]These data were coded such that only "Definitely yes" was coded as "yes." All other responses were coded as "no."
[2]These data were coded such that only "yes" responses to the DC CV who also rated their response a 10 (very certain) were coded as "yes." All other responses were coded as "no."
*Data not available.

Table 4: Logistic Model Using Actual Cash Donation Data (n=208)

Explanatory Variables	Estimated Coefficient (Standard Error)
Constant	.3497 (3.7921)
Offer amount	-.07222* (.0248)
My decision about whether or not to pay was based on the number of feet of road that would be removed if I agreed to pay. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-2.0476* (.4579)
When I was deciding whether to pay, I considered the fact that I already pay for environmental projects through taxes. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	.4739 (.3510)
I would rather see the money go to a better project. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	1.8008* (.4705)
It is important to me that the area is designated "wilderness" after the roads are closed. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-1.7346* (.6550)
How likely is it you will visit Grand Canyon National Park in the future? 1=Very unlikely, 2=Somewhat unlikely, 3=Somewhat likely, 4=Very likely, 5=Not sure	.2140 (.3453)
All areas of National Parks should be easily accessible by roads. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-.3566 (.3132)
It is important to me that future generations be able to enjoy wilderness areas. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-1.4094 (.9562)
I would like for wilderness areas to be preserved even if I never get to visit them. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	1.0741 (.9581)
I think it is everyone's responsibility to help the environment any way we can. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	.6042 (.6584)

Table 4: Logistic Model Using Actual Cash Donation Data (n=208) (continued)

Explanatory Variables	Estimated Coefficient (Standard Error)
I care about wilderness areas outside of Wisconsin. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-.6799 (.6692)
National parks should be managed to preserve native species. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	-.9154 (.7561)
National parks should be managed to preserve wilderness areas. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	.8394 (.9393)
My responses to this study will be important in future decisions about the environment. 1=Definitely True, 2=Somewhat True, 3=Somewhat False, 4=Definitely False	.8508* (.4124)
Which category best describes where you currently live? 1=Urban, 2=Suburban, 3=Rural	-.8366* (.3734)
What is the highest grade or year of school you have completed? 1=Eighth grade or less, 2=Some high school, 3=High school graduate, 4=Some college or technical school, 5=Technical or trade school graduate, 6=College graduate, 7=Some graduate work, 8=Advanced Degree	.1507 (.2115)
What was your total household income before taxes and deductions? 1=Less than $10,000, 2=$10,000 to $19,999, 3=$20,000 to $29,999, 4=$30,000 to $39,999, 5=$40,000 to $49,999, 6=$50,000 to $59,999, 7=$60,000 to $69,999, 8=$70,000 to $79,999, 9=$80,000 to $99,999, 10=$100,000 or more	.4768* (.1619)
-2*Log Likelihood	-82.538
Percentage of Yes Responses Predicted Correctly	81%
Percentage of No Responses Predicted Correctly	97%

*Significantly different from zero at 5% significance level.

Table 5: Observed and Predicted Responses to WTP Question for DC Respondents (N=458)

Observed Response to DC CV Question	Predicted Response to DC Actual Cash Donation Question	
	Yes	**No**
Yes	95	108*
No	9*	246

*Indicates inconsistent responses.

Table 6: Percentage Inconsistent Responses by Offer Amount

Offer Amount	Inconsistent Responses as a Percent of Yes Responses to DC CV Question	Inconsistent Responses as a Percent of both Yes and No Responses to DC CV Question
$1	30%	18%
$5	30%	18%
$8	54%	29%
$12	46%	27%
$15	44%	23%
$50	79%	22%
$75	100%	44%
$100	100%	24%
$150	100%	21%
$200	100%	27%

Figure 1: Estimated Logit Distributions
for Follow-up Certainty Question Data

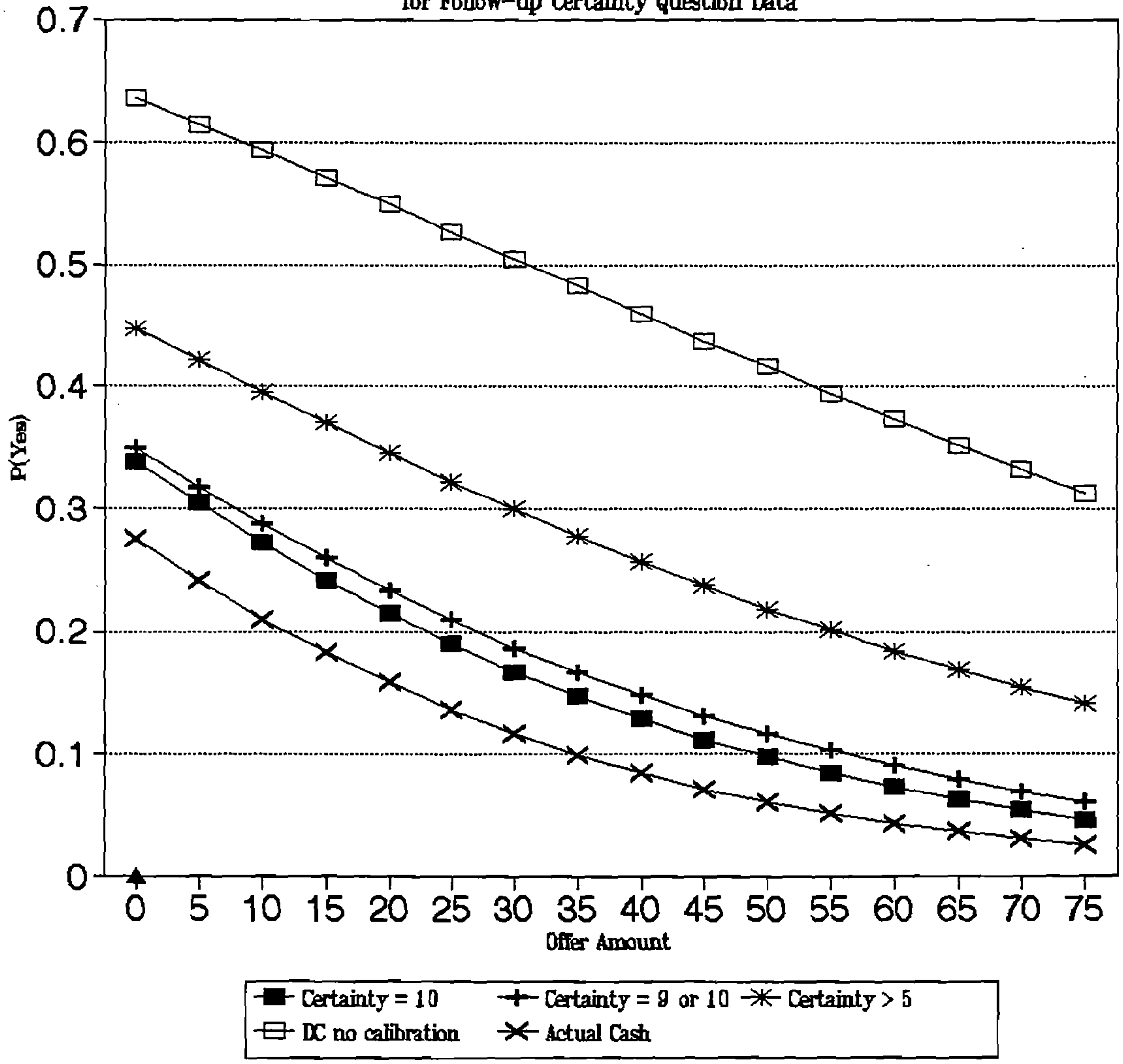
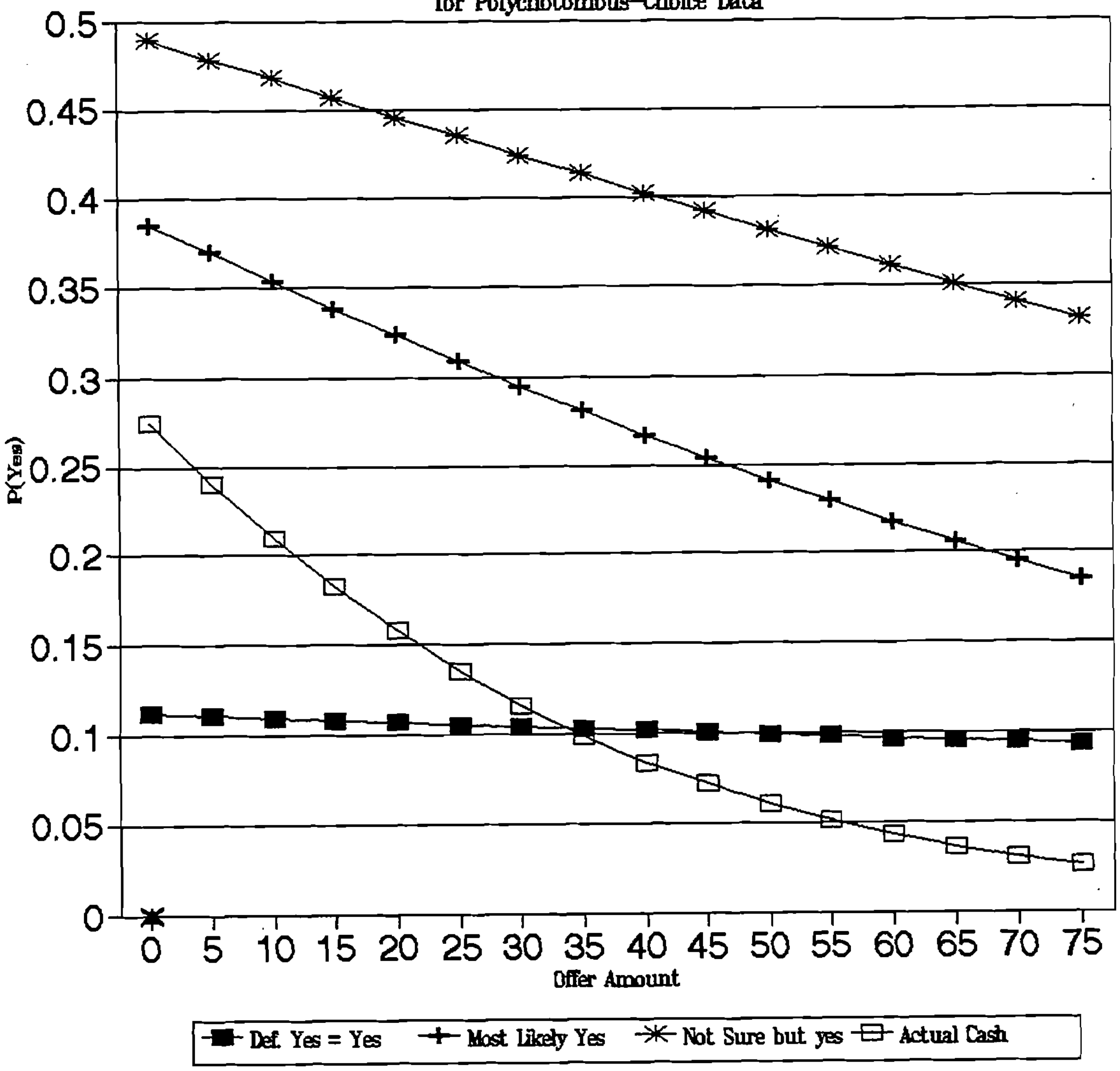


Figure 2: Estimated Logit Distributions for Polychotomous–Choice Data



TESTING THE CONSISTENCY OF NESTED LOGIT MODELS WITH UTILITY MAXIMIZATION*

by

JOSEPH A. HERRIGES AND CATHERINE L. KLING
Department of Economics, Iowa State University

ABSTRACT

The Nested Multinomial Logit (NMNL) model is used extensively in modeling consumer choices among discrete alternatives when the number of alternatives is large. Unfortunately, applied researchers often find that estimated NMNL models fail to meet the Daly-Zachary-McFadden (DZM) sufficient conditions for consistency with stochastic utility maximization. Börsch-Supan (1990) provides a relaxed set of conditions to test for consistency. While these conditions are increasingly cited, they are seldom tested. This paper corrects and extends Börsch-Supan's Theorem 2, providing simple necessary conditions on first, second, and third derivatives of choice probabilities and a graph of the bounds they place on dissimilarity parameters.

March 29, 1995

* The authors would likely acknowledge financial support provided by Iowa State University's Agricultural Experiment Station and the W-133 Western Regional Project.

1. Introduction

The Nested Multinomial Logit (NMNL) model is used extensively in modeling consumer choices among discrete alternatives when the number of alternatives is large. Prominent examples can be found in empirical studies of transportation mode and travel demand [e.g., Domencich and McFadden (1975), Ben-Akiva and Lerman (1985), and Train (1986)], housing choice [e.g., Börsch-Supan (1986,1987)], and recreational site selection [e.g., Hausman, Leonard, and McFadden (1992) and Morey, Rowe, and Watson (1993)]. The popularity of NMNL stems largely from its compromise position between the traditional Multinomial Logit (MNL) and Multinomial Probit (MNP) specifications. On the one hand, both MNL and NMNL models yield closed-form choice probabilities, greatly simplifying the estimation process by avoiding the numerical or Monte Carlo integration techniques required for MNP.[1] On the other hand, MNL severely restricts the correlation patterns among choice alternatives, imposing the well-known assumption of the independence of irrelevant alternatives (IIA).[2] Nested logit relaxes this assumption, organizing like alternatives into groups and allowing different correlation patterns between groups than within groups. While NMNL imposes more structure than its probit counterpart, considerable flexibility is gained over MNL.

Another feature of the nested logit specification that is often cited in the literature is that, under certain conditions, NMNL is consistent with stochastic utility maximization [McFadden (1981)]. The sufficient conditions for consistency (i.e., the Daly-Zachary-McFadden (DZM) conditions in Börsch-Supan (1990)) require the nested logit's dissimilarity coefficients to lie within the unit interval. This condition, in turn, ensures that the density function will be non-

---

[1] Traditionally, MNP has been viewed as practical only for choice problems involving fewer than five alternatives [Maddala (1983)]. Recent developments in econometric methods [e.g., McFadden (1989) and Börsch-Supan and Hajivassiliou (1993)] suggest, however, that MNP may now be feasible for problems involving more alternatives.

[2] See McFadden, Tye, and Train (1977) and McFadden (1981)

negative. Unfortunately, applied researchers often find that estimated NMNL models fail to meet the DZM conditions [Jones and Stokes (1987), Hausman, Leonard, and McFadden (1992), and Train, Ben-Akiva, and Atherton (1989)]. Börsch-Supan (1990) has recently suggested that these failures are due, in part, to the DZM conditions being too stringent. He argues that, just as flexible functional forms used in demand analysis are viewed as approximations to the true underlying demand system, so too should the nested logit specification be viewed as an approximation. As a result, stochastic utility maximization should not be expected to hold globally, but only within the region of "...data points that are sensible for a specific application of the choice model..." [Börsch-Supan (1990, p. 377)] Börsch-Supan develops a relaxed set of conditions to test for consistency.

The purpose of this short paper is two-fold. First, while the Börsch-Supan (BS) conditions are increasingly being cited in the literature as an alternative to the DZM conditions [e.g., Cameron (1989), Hensher (1986), and Morey (1994)], they have yet to be explicitly tested. This is in part due to the lack of explicit formulae for the conditions when numerous choice alternatives exist. This paper corrects and extends Börsch-Supan's Theorem 2, providing simple necessary conditions on the first, second, and third derivatives of choice probabilities. Second, we examine the extent to which the BS conditions are likely to relax the DZM conditions. We find that, for applications with several alternative groups, the BS conditions do not expand the acceptable range for the dissimilarity coefficients far beyond the unit interval.

2. The nested multinomial logit model

Following the notation in Börsch-Supan (1990), let I denote the total number of discrete alternatives from which the consumer can choose and T denote the number of consumers. In an application to recreational demand, the alternatives might include various fishing and boating sites within a region. Consumer t is assumed to receive utility u_{it} from the selection of alternative i, with

$$u_{it} = v_{it} + \varepsilon_{it} \tag{1}$$

where v_{it} denotes the deterministic component of individual utility and ε_{it} denotes the random disturbance.[3]

The nested-logit model results when the disturbance vector $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tI})$ is assumed to be i.i.d. and drawn from a generalized extreme value (GEV) distribution [McFadden (1978)]. The alternatives are organized into K groups of similar alternatives, with J(k) indexing the first alternative within the kth group and I(k) denoting the number of alternatives within the kth group.[4] Within the recreational demand literature, for example, recreational fishing sites might be grouped into shore, pier, and boating alternatives. Given the assigned groupings and the GEV distributional assumption, the probability that an individual will select any specific alternative i is then given by:[5]

$$P_i = P_{i|g(i)} Q_{g(i)} \tag{2}$$

where g(i) denotes the group to which alternative i belongs,

$$P_{i|g(i)} = exp(v_i / \theta_{g(i)}) / E[g(i)] \tag{3}$$

---

[3] Typically, the deterministic component is modeled as a function of individual and alternative characteristics (X_{it}) (i.e., $v_{it} = f(X_{it}; \beta_t)$, with f often restricted to being linear in the X_{it}'s). The random component is assumed to capture inter- and intra-personal variations in tastes. See McFadden (1981) for additional discussion of the stochastic utility maximization hypothesis.

[4] As in Börsch-Supan (1990), we limit our attention to two-level nested logit models. Although additional nesting levels can be employed, the vast majority of applications in the literature are two-level models.

[5] The subscript t is dropped throughout the remainder of the paper in order to simplify the notation.

denotes the conditional probability of selecting alternative i given group g(i) has been selected, with

$$E(k) \equiv \sum_{n=J(k)}^{J(k)+I(k)-1} exp(v_n / \theta_k), \tag{4}$$

and

$$Q_k \equiv \frac{E(k)^{\theta_k}}{\sum_{k=1}^{K} E(k)^{\theta_k}} \tag{5}$$

denotes the marginal probability that any alternative from within group k is selected. The parameter θ_k is the so-called dissimilarity parameter for group k. The nested logit model reduces to the multinomial logit model if $\theta_k = 1 \; \forall \; k = 1,\ldots,K$.

3. Consistency conditions

McFadden (1981) establishes the conditions under which a set of choice probabilities (i.e., P_i's) will be consistent with stochastic utility maximization. As noted in Börsch-Supan (1990, p. 375), these conditions include:

C.1 $$P_i(v) > 0, \qquad \sum_{i=1}^{I} P_i(v) = 1, \qquad P_i(v) = P_i(v+\alpha) \quad \forall \alpha \in R, \tag{4}$$

where $v \equiv (v_1,\ldots,v_I)$, and

C.2 $$\partial P_i(v) / \partial v_j = \partial P_j(v) / \partial v_i \,. \tag{5}$$

In addition,

C.3 P_i must have nonnegative even and nonpositive odd mixed partials derivatives with respect to components of v other than v_i.[6]

This last condition ensures that the implied probability density function will be nonnegative.

It is straightforward to verify that the NMNL model automatically satisfies the first two compatibility conditions. However, in order for condition C.3 to be satisfied globally (i.e., $\forall\, v \in R^I$), the dissimilarity coefficients are restricted to lie within the unit interval [McFadden (1979), Daly and Zachary (1979)]; i.e.,

$$0 < \theta_i \leq 1 \quad \forall\, i. \tag{6}$$

The primary contribution of Börsch-Supan (1990) was to note that, while the DZM condition is indeed required for global consistency, this condition is too restrictive if the nested logit model is viewed as a local approximation. Instead, condition C.3 should be applied only for that subset of R^I (i.e., $A \subseteq R^I$) in which relevant deterministic components, v, are likely to lie.[7] The author's Theorem 1 provides a formal proof of this proposition for any set of choice probabilities. Börsch-Supan's Theorem 3 then establishes that, for a two-level nested logit, condition C.3 results in nonnegativity restrictions that are signed by polynomials in the Q_k's.

While Theorems 1 and 3 provide the theoretical foundation for Börsch-Supan's relaxation of the DZM condition, it is the author's Theorem 2 that provides a practical translation of this restriction to the two-level nested logit model. The theorem notes that condition C.3 requires

[6] Börsch-Supan (1990, p. 375, eq. 6) incorrectly lists this condition as requiring P_i to have nonnegative mixed partial derivatives with respect to components of v. However, McFadden's (1981, p. 211) condition SS 5.4 ensures that P_i with have nonnegative mixed partial derivatives *with respect to* q_i (the cost of alternative i). Since $\partial v_i/\partial q_i \leq 0$, the mixed partial derivatives of P_i with respect to v_i must alternate in sign, beginning with a nonpositive sign.

[7] Börsch-Supan (1990, fn. 4) defines this relevant region to be comprised of the data points for all observed and projected deterministic utility components.

$$Q_k \geq (1-\theta_k)/\theta_k, \qquad k = 1,\ldots,K. \tag{7}$$

The heart of our contribution lies in correcting and extending the results of Theorem 2 and examining the extent to which it is likely to expand the set of consistent NMNL models. Specifically, we have:

Theorem 1. In two-level NMNL models, the following are necessary conditions for consistency with stochastic utility maximization:

$$Q_k \geq \tau_k, \qquad k = 1,\ldots,K. \tag{8}$$

$$2(\tau_k - Q_k)^2 + \tau_k Q_k \geq \tau_k \qquad \forall\, k \in G_3 \equiv \{h | I(h) \geq 3\} \tag{9}$$

and

$$6(Q_k - \tau_k)^3 + \tau_k [2(Q_k - 1) - \tau_k](1 - Q_k) \geq 0 \quad \forall\, k \in G_4 \equiv \{h | I(h) \geq 4\} \tag{10}$$

where $\tau_k \equiv (\theta_k - 1)/\theta_k$.

The proof of Theorem 1 follows by simple, though tedious, differentiation of equation (2) and is provided in Appendix A. Equations (8), (9), and (10) correspond to restrictions implied by C.3 for the first, second, and third mixed partial derivatives of P_i, with equation (8) correcting the sign error in Börsch-Supan's (1990) Theorem 2, equation (25).

The results of Theorem 1 place implicit restrictions on the dissimilarity coefficients, θ_k. The restrictions corresponding to equation (8) and (9) are made explicit in the following corollary:[8]

Corollary 1. In two-level NMNL models, consistency with stochastic utility maximization places the following necessary restrictions on dissimilarity coefficients:

[8] While the explicit restriction on θ_k implied by equation (10) can be derived, it is lengthy and not presented here. The left-hand side of equation (10) has three roots, only one of which is real.

$$\theta_k \le \frac{1}{1-Q_k}, \qquad k=1,\ldots,K. \tag{11}$$

and

$$\theta_k \le \frac{4}{3(1-Q_k)+\sqrt{(1+7Q_k)(1-Q_k)}} \qquad \forall\, k \in G_3 \equiv \{h \mid I(h) \ge 3\} \;. \tag{12}$$

Proof. Corollary 1 follows from Theorem 1 by explicitly solving for the θ_i's in equations (8) and (9). Q.E.D.

There are several things to note about the results of Theorem 1 and its corollary. First, the restrictions imposed on θ_k by consistency condition C.3 are expressed in terms of Q_k, with no cross-group terms involved. As seen below, this makes it straightforward to solve for and check the consistency conditions. Second, for groups with three or more alternatives, the inequality condition in equation (12) will always be more restrictive than that in equation (11), since

$$\frac{4}{3(1-Q_k)+\sqrt{(1+7Q_k)(1-Q_k)}} \le \frac{1}{1-Q_k} \;. \tag{13}$$

Similarly, the third order partial derivative restrictions implied by equation (10) dominate the second order partial restrictions in equation (12) for groups with four or more alternatives.

The conditions in Theorem 1, together with those identified in C.1 and C.2, provide a complete set of the necessary and sufficient conditions for consistency with stochastic utility maximization when there are four or fewer alternatives per nest. While Theorem 1 does not provide a complete enumeration of the conditions required for local consistency in models with more than four alternatives per choice set, it does contain a set of readily verified necessary conditions for the NMNL that can either be tested ex post or imposed in the estimation process.

The restrictions in Theorem 1 can also be used to examine the extent to which Börsch-Supan's approach is likely to expand the set of NMNL models that are consistent with utility theory. Table 1, using the results of Theorem 1 and Corollary 1, lists the admissible upper bounds for θ_k.[9] For example, for a group selected roughly half of the time by consumers (i.e., Q_k = .5), the first derivative restrictions in equation (11) restricts the corresponding dissimilarity coefficient to lie below 2.00. This suggests considerable flexibility in θ_k's range when compared to the upper bound of 1.00 in the global DZM conditions. Unfortunately, the second and third order derivative conditions narrow these gains considerably, requiring θ_k to lie below 1.28 once the implicit restrictions in equation (10) are imposed. For groups with lower marginal choice probabilities, the gains over DZM are even smaller. When Q_k reaches .25, for example, θ_k is restricted to lie between zero and 1.05, an expansion of only five percent in the acceptable region.

Figure 1 illustrates these results. The shaded region indicates the bounds place on θ_k for a given Q_k by the DZM consistency conditions. The upper limits on θ_k implied by Theorem 1 and identified in Corollary 1 are illustrated as well. As note above, the first, second, and third order conditions are progressively more restrictive, providing little additional range for θ_k when Q_k lies below .5.

4. Conclusions

Börsch-Supan (1990) argued that the traditional DZM conditions for consistency with stochastic utility maximization were too stringent, requiring global conformity with utility theory. Instead, he suggested that a local approximation perspective be adopted, imposing consistency only within the relevant range of marginal choice probabilities. In this paper, we

[9] MathCad 5.0 Plus was used to solve for the roots of equation (10) and to verify that the resulting real root provided an upper bound on θ_k.

have corrected and expanded the necessary conditions provided by Börsch-Supan's Theorem 2, simplifying the process of testing or imposing these local restrictions. In addition, an examination of the resulting conditions reveals that, while a local approximation perspective does allow the dissimilarity coefficient to lie outside of the unit interval, the additional maneuvering room that it provides applied economists is small when several groups are included in the NMNL model.

5. Appendix A: Proof of Theorem 1

The proof of Theorem 1 follows directly from differentiating equation (2). Equations (8), (9), and (10) of the theorem correspond to applying condition C.3 for all first, second, and third order partial derivatives of P_i. Let

$$R \equiv \sum_{m=1}^{K} E(m)^{\theta_m}. \tag{A.1}$$

Then equation (2) can be rewritten as:

$$P_i(v) = exp(v_i / \theta_{g(i)}) E(k)^{\theta_{g(i)}-1} R^{-1} \tag{A.2}$$

Using

$$\delta_{ij} = \begin{cases} 1 & g(i) = g(j) \\ 0 & g(i) \neq g(j) \end{cases} \tag{A.3}$$

to indicate that two alternatives are in the same subgroup, we have

$$\partial P_i / \partial v_j = \delta_{ij}(\theta_{g(j)} - 1)\{P_i / E[g(j)]\} exp(v_j / \theta_{g(j)}) / \theta_{g(j)} - [P_i / R] E[g(j)]^{\theta_{g(j)}-1} exp(v_j / \theta_{g(j)})$$

$$= P_i \left[\delta_{ij} \frac{[\theta_{g(j)} - 1]}{\theta_{g(j)}} P_{j|g(j)} - P_j \right]$$

$$= P_i P_j A_{ij} \tag{A.4}$$

where

$$A_{ij} \equiv \left[\frac{\delta_{ij} \tau_{g(j)}}{Q_{g(j)}} - 1 \right] \tag{A.5}$$

In order for the first partial derivatives of P_i to have the nonpositive sign required by condition C.3, then A_{ij} must be non-positive, yielding the condition in equation (8) of Theorem 1.

In setting up the second derivative equations, it is useful to note that:

$$\begin{aligned}
\frac{\partial A_{ij}}{\partial v_k} &\equiv - \frac{\delta_{ij} \tau_{g(j)}}{Q^2_{g(j)}} \frac{\partial Q_{g(j)}}{\partial v_k} \\
&= - \frac{\delta_{ij} \tau_{g(j)}}{Q^2_{g(j)}} \left[\delta_{jk} \frac{E^{\theta_{g(k)}-1} e^{v_k/\theta_{g(k)}}}{R} - \frac{E^{\theta_{g(j)}} E^{\theta_{g(k)}-1} e^{v_k/\theta_{g(k)}}}{R^2} \right] \\
&= - \frac{\delta_{ij} \tau_{g(j)}}{Q^2_{g(j)}} \left[\delta_{jk} P_k - P_k Q_{g(j)} \right] \\
&= - \frac{\delta_{ij} P_k \tau_{g(j)}}{Q^2_{g(j)}} \left[\delta_{jk} - Q_{g(j)} \right]
\end{aligned} \tag{A.6}$$

The second derivatives follow from equations (A.4) and (A.6) as:

$$\frac{\partial^2 P_i}{\partial v_k \partial v_j} = \frac{\partial}{\partial v_k}\left(P_i P_j A_{ij}\right)$$

$$= P_j A_{ij}\left(\frac{\partial P_i}{\partial v_k}\right) + P_i A_{ij}\left(\frac{\partial P_j}{\partial v_k}\right) + P_i P_j\left(\frac{\partial A_{ij}}{\partial v_k}\right)$$

$$= P_i P_j P_k A_{ij} A_{ik} + P_i P_j P_k A_{ij} A_{jk} - P_i P_j P_k\left(\frac{\delta_{ij}\tau_{g(j)}}{Q_{g(j)}^2}\left[\delta_{jk} - Q_{g(j)}\right]\right) \tag{A.7}$$

$$= P_i P_j P_k\left[A_{ij} A_{ik} + A_{ij} A_{jk} + \left(\frac{\delta_{ij}\tau_{g(j)}}{Q_{g(j)}^2}\left[Q_{g(j)} - \delta_{jk}\right]\right)\right]$$

$$= P_i P_j P_k\left[A_{ij} A_{ik} + A_{ij} A_{jk} + \frac{\delta_{ij}\tau_{g(j)}}{Q_{g(j)}} - \frac{\delta_{ij}\delta_{jk}\tau_{g(j)}}{Q_{g(j)}^2}\right]$$

Condition C.3 requires that these second derivatives be nonnegative. It is clear from a quick perusal of equation (14) that the second order conditions add nothing to the first order conditions *unless* alternatives i, j, and k are in the same group. Since the first order conditions require that $A_{ij} \le 0 \; \forall i, j$, the first two terms in the square brackets are nonnegative. The third term is always positive, leaving only the fourth term, which is negative iff $\delta_{ij} = \delta_{jk} = 1$ (i.e., the alternatives are all in the same group). Thus, the second order conditions add the following constraint for groups with three or more alternatives:

$$\left[A_{ij} A_{ik} + A_{ij} A_{jk} + \frac{\delta_{ij}\tau_{g(j)}}{Q_{g(j)}} - \frac{\delta_{ij}\delta_{jk}\tau_{g(j)}}{Q_{g(j)}^2}\right] \ge 0 \tag{A.8}$$

$\Rightarrow$

$$A_{ij} A_{ik} Q_{g(j)}^2 + A_{ij} A_{jk} Q_{g(j)}^2 + \tau_{g(j)} Q_{g(j)} \ge \tau_{g(j)} \tag{A.9}$$

Rearranging equation (A.9) yields equation (9) from Theorem 1.

Finally, let:

$$B_{ijk} \equiv \left[A_{ij}A_{ik} + A_{ij}A_{jk} + \frac{\delta_{ij}\tau_{g(j)}}{Q_{g(j)}} - \frac{\delta_{ij}\delta_{jk}\tau_{g(j)}}{Q^2_{g(j)}} \right] \tag{A.10}$$

Using equation (A.7), we then have:

$$\begin{aligned}
\frac{\partial^3 P_i}{\partial v_l \partial v_k \partial v_j} &= \frac{\partial}{\partial v_l}\left(P_i P_j P_k B_{ijk}\right) \\
&= P_j P_k B_{ijk}\left(\frac{\partial P_i}{\partial v_l}\right) + P_i P_k B_{ijk}\left(\frac{\partial P_j}{\partial v_l}\right) + P_i P_j B_{ijk}\left(\frac{\partial P_k}{\partial v_l}\right) + P_i P_j P_k\left(\frac{\partial B_{ijk}}{\partial v_l}\right) \\
&= P_i P_j P_k P_l B_{ijk} A_{il} + P_i P_j P_k P_l B_{ijk} A_{jl} + P_i P_j P_k P_l B_{ijk} A_{kl} + P_i P_j P_k P_l\left(\frac{\partial B_{ijk}}{\partial v_l} P_l^{-1}\right) \\
&= P_i P_j P_k P_l\left[B_{ijk}\left(A_{il} + A_{jl} + A_{kl}\right) + P_l^{-1}\left(\frac{\partial B_{ijk}}{\partial v_l}\right)\right]
\end{aligned} \tag{A.11}$$

Condition C.3 for consistency with stochastic utility maximization requires that the third mixed partial derivatives be nonpositive, so that:

$$B_{ijk}\left(A_{il} + A_{jl} + A_{kl}\right) + P_l^{-1}\left(\frac{\partial B_{ijk}}{\partial v_l}\right) \le 0 \tag{A.12}$$

The first thing to note about equation (A.12) is that the third derivative adds additional constraints only if all four alternatives are in the same group (i.e., $\delta_{ij} = \delta_{jk} = \delta_{lk} = 1$). The argument is as follows. Since the second order conditions require that $B_{ijk} \ge 0$ and the first order conditions require that $A_{ij} \le 0 \; \forall i, j$, the first term on the LHS of equation (A.12) is nonpositive. Since B_{ijk} is a function of $Q_{g(j)}$ and τ_j only, then $\partial B_{ijk} / \partial v_l = 0$ unless $\delta_{ij} = \delta_{jk} = \delta_{lk} = 1$ and condition (A.12) will always hold as long as the first and second derivative conditions hold.

Turning to the remaining case, we begin by noting that, for $\delta_{ij} = \delta_{jk} = \delta_{lk} = 1$:

$$\frac{\partial B_{ijk}}{\partial v_l} = = \frac{P_l}{Q^3_{g(i)}} \left\{4\tau_{g(i)}(\tau_{g(i)} - Q_{g(i)}) - \tau_{g(i)}\left(Q_{g(i)} - 2\right)\right\}\left(1 - Q_{g(i)}\right) \tag{A.13}$$

Condition (A.12) becomes:

$$B_{ijk}\left(A_{il} + A_{jl} + A_{kl}\right)Q^3_{g(i)} + \left(\left\{4\tau_{g(i)}(\tau_{g(i)} - Q_{g(i)}) - \tau_{g(i)}\left(Q_{g(i)} - 2\right)\right\}\left(1 - Q_{g(i)}\right)\right) \le 0 \tag{A.14}$$

Expanding the left-hand side of equation (A.14) and collecting terms yields equation (10) of Theorem 1. Q.E.D.

References

Ben-Akiva, M., and S. R. Lerman, 1985, Discrete choice analysis: theory and application to travel demand (M. I. T. Press, Cambridge, Massachusetts).

Börsch-Supan, A. 1986, Household formation, housing prices, and public policy impacts, Journal of Public Economics, 30, 145-164.

Börsch-Supan, A., 1987, Economic analysis of discrete choice (Springer-Verlag, Berlin).

Börsch-Supan, A., 1990, On the compatibility of nested logit models with utility maximization, Journal of Econometrics, 43, 373-388.

Börsch-Supan, A., and V. A. Hajivassiliou, 1993, Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models, Journal of Econometrics, 58, 347-368.

Cameron, T., 1989, Econometric analysis of discrete choice - book review, Journal of Economic Literature, 27, 663-664.

Daly, A., and S. Zachary, 1979, Improved multiple choice models, in: David Hensher and Q. Dalvi, eds., Identifying and measuring the determinants of mode choice (Teakfield, London) 335-357.

Domencich, T., and D. McFadden, 1975, Urban travel demand: a behavioral analysis (North Holland: Amsterdam).

Hausman, J. A., G. K. Leonard, and D. McFadden, 1992, A utility-consistent, combined discrete choice and count data model: Assessing recreational use losses due to natural resource damage, unpublished mimeograph, March.

Hensher, D. A., 1986, Sequential and full information maximum likelihood estimation of a nested logit model, Review of Economics and Statistics, 657-667.

Jones and Stokes Associates, 1987, Southern Alaska sport fishing economic study, prepared for Alaska Department of Fish and Game, November.

Maddala, G. S., 1983, Limited-dependent and quantitative variables in econometrics (Cambridge University Press, Cambridge).

McFadden, D., 1978, Modeling the choice of residential location, in: Karquist, et al., eds., Spatial interaction theory and planning models (North-Holland Publishing Company: Amsterdam) 75-96.

McFadden, D., 1979, Quantitative methods for analyzing travel behavior of individuals: Some recent developments, in: David Hensher and P. Stopher, eds., Behavioral travel modeling (Croom Helm, London), 279-318.

McFadden, D., 1981, Econometric models of probabilistic choice, in: Charles F. Manski and Daniel McFadden, eds., Structural analysis of discrete data with econometric application

McFadden, D., 1989, A method of simulated moments for estimation of discrete response models without numerical integration, Econometrica, 57, 995-1026.

McFadden, D., W. Tye, and K. Train, 1977, An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model, Transportation Research Record, 637, 39-45.

Morey, E., 1994, Two RUMS uncloaked: Nested-logit models of site choice and nested-logit models of participation and site choice, discussion paper, Department of Economics, University of Colorado, Boulder, February.

Morey, E. R., R. D. Rowe, and M. Watson, 1993, A repeated nested-logit model of Atlantic salmon fishing, American Journal of Agricultural Economics, 75, 578-592.

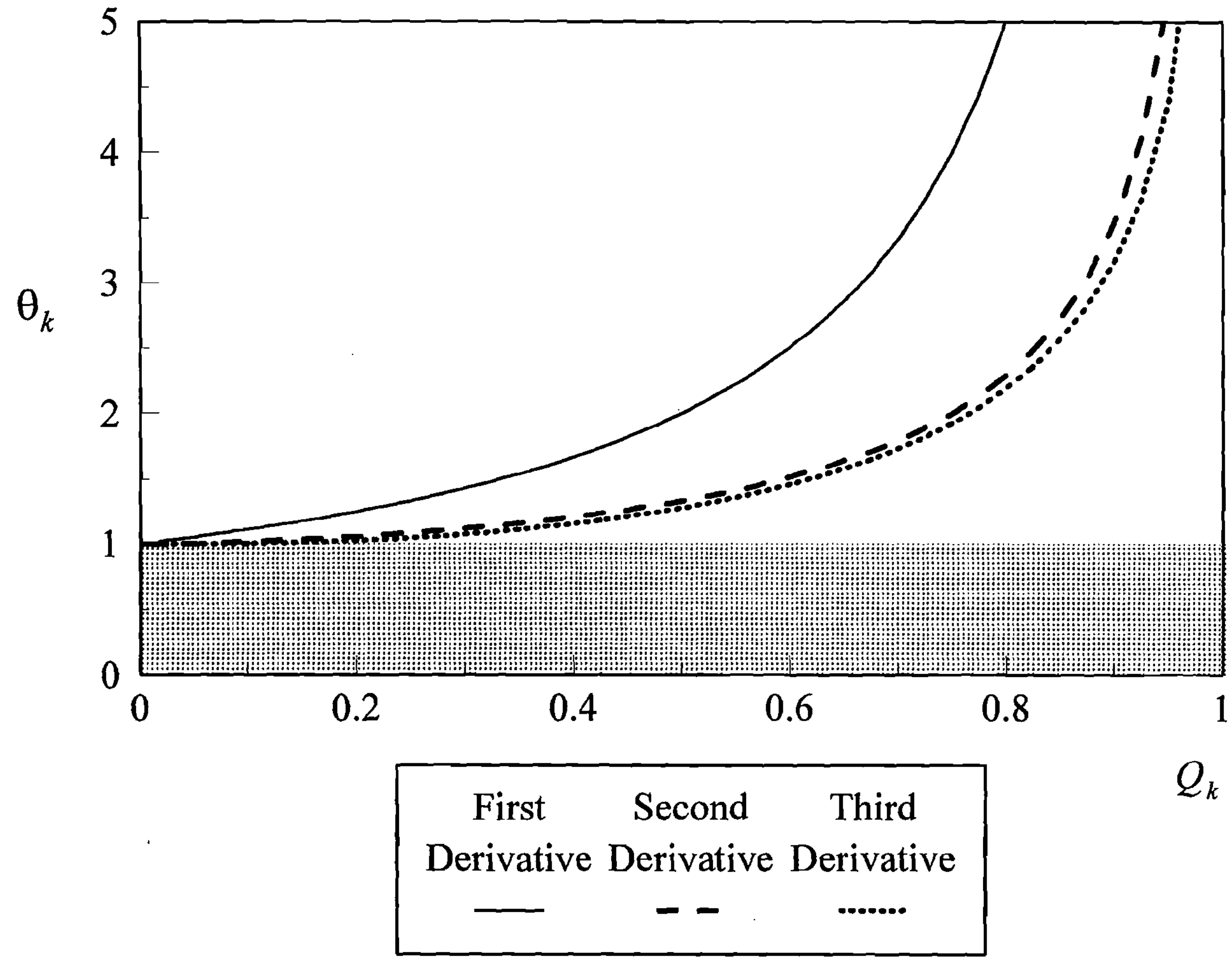
Train, K., 1986, Qualitative choice analysis: theory, econometrics, and an application to automobile demand (M. I. T. Press, Cambridge, Massachusetts).

Train, K. E., M. Ben-Akiva, and T. Atherton, 1989, Consumption patterns and self-selecting tariffs, The Review of Economics and Statistics, 71, 62-73.

Table 1
Upper Limits on θ_k

	Derivative Restrictions		
Q_k	First	Second	Third
0	1.00	1.00	1.00
.05	1.05	1.00	1.00
.10	1.11	1.02	1.00
.15	1.18	1.03	1.01
.20	1.25	1.06	1.03
.25	1.33	1.09	1.05
.30	1.43	1.12	1.08
.35	1.54	1.16	1.11
.40	1.67	1.21	1.16
.45	1.82	1.27	1.21
.50	2.00	1.33	1.28
.55	2.22	1.41	1.36
.60	2.50	1.52	1.46
.65	2.86	1.64	1.58
.70	3.33	1.79	1.73
.75	4.00	2.00	1.92
.80	5.00	2.29	2.19
.85	6.67	2.72	2.56
.90	10.00	3.47	3.16
.95	20.00	5.21	4.39

Figure 1
Limits on Theta



Effects of Data Conditioning, Sample Design, and Aggregation on Random Utility Model Estimates: Some Monte Carlo Results

Steven M. Waters
F. Reed Johnson
Robert B. Fowler

of

Triangle Economic Research
Durham, NC

Effects of Data Conditioning, Sample Design, and Aggregation on Random Utility Model Estimates: Some Monte Carlo Results

Steven M Waters, F. Reed Johnson, and Robert B. Fowler
Triangle Economic Research

Introduction

Recreation demand researchers are increasingly using travel cost models based on the random utility model.[1] Several examples of this trend include the articles by Parsons and Kealy (1992), Morey, Shaw and Rowe (1991), and Kaoru, Smith and Liu (1995). Economists generally enter the RUM arena with statistical intuition derived from the ordinary least squares (OLS) regression model. However, RUM econometric results often run counter to this intuition, and we are left to wonder if the unexpected outcome is a result of ill-conditioned data, poor specification, or simply an artifact of the nonlinearity of the model itself. This paper explores the consequences of various data-conditioning and model-specification problems by generating pseudo-data from a known distribution and assigning choices based on probability functions of known parameters.

Rum applications often have to cope with various practical problems involving data availability and quality. For example, in a recreation demand setting the RUM allows the modeling of potentially numerous substitute sites. This is an advantage over the traditional travel cost model. However, the advantage of the RUM comes at the cost of having to collect more

[1] See Ben-Akiva and Lerman (1985) and Train (1986) for a complete presentation of the random utility model and discrete-choice analysis.

information since the researcher must have information not only on the site that the household visited, but also on the sites that the household considered when making the decision. Identifying the sites a household considers requires the researcher to specify households' choice set. In the absence of detailed information on recreators' decision calculus, the researcher often resorts to ad hoc rules for identifying choice sets and site-characteristic variables. The success the researcher has in specifying the relevant choice set and in identifying appropriate site characteristics for each site may affect the quality of the demand estimation. This paper reports the results of three experiments to evaluate the consequences of various choice-set and site-characteristic modeling strategies.

This paper begins with an analysis of errors-in-variables. The purpose of this experiment is to examine the results of estimating a model with a single poorly measured explanatory variable. This experiment indicates that within the confines of the assumed parametric distribution there is an observable bias in all parameters estimated in the model, and not just on the ill-conditioned variable. However, the bias appears to be relatively mild for the particular parameters assumed in our experiment.

The second experiment examines sample design. This experiment involves a sample design that oversamples particular sites -- a design that is not uncommon in recreation demand studies. This experiment shows that neglecting weights strongly effects both model estimates and willingness-to-pay values.

The final experiment simulates having limited site-characteristic information. In this experiment we demonstrate the effects of a particular strategy for incorporating trips to sites into the analysis for which few or no site characteristics exist. Provided the sites with limited information are chosen randomly, this experiment suggests that there are small efficiency losses in not collecting consistent site characteristics across all sites used in the model.

Model and Data Generation

Each of the experiments estimates a two-level nested logit model. The nested logit framework has been suggested as a realistic model of decision making and mitigates the restrictive assumption of Independence of Irrelevant Alternatives inherent in the logistic functional form.[2] The nested logit models presented here are completely general, but to facilitate the discussion we will discuss them in the context of recreation demand.

The decision structure of the model is shown graphically in Figure 1. The first decision in this framework is to choose between two activities. Given a particular activity, the second decision is what site to visit. Within each activity, the site choice is modeled with two explanatory variables which we have denoted as site characteristic n and distance. If activity A is fishing, then site characteristic 1 could be the stock level of fish in the river or lake. Similarly, if activity B is hiking, then site characteristic 2 could be the length of hiking trails at the site. In both activities, the distance variable is generated from a distribution similar to an actual empirical data set. The limited site information studies employ a slightly different setup which will be explained below.

[2] See the independence of irrelevant alternatives (IIA) discussion in Ben-Akiva and Lerman (1985). Morey (1994) shows why a nested logit model mitigates, but does not eliminate, the IIA problem.

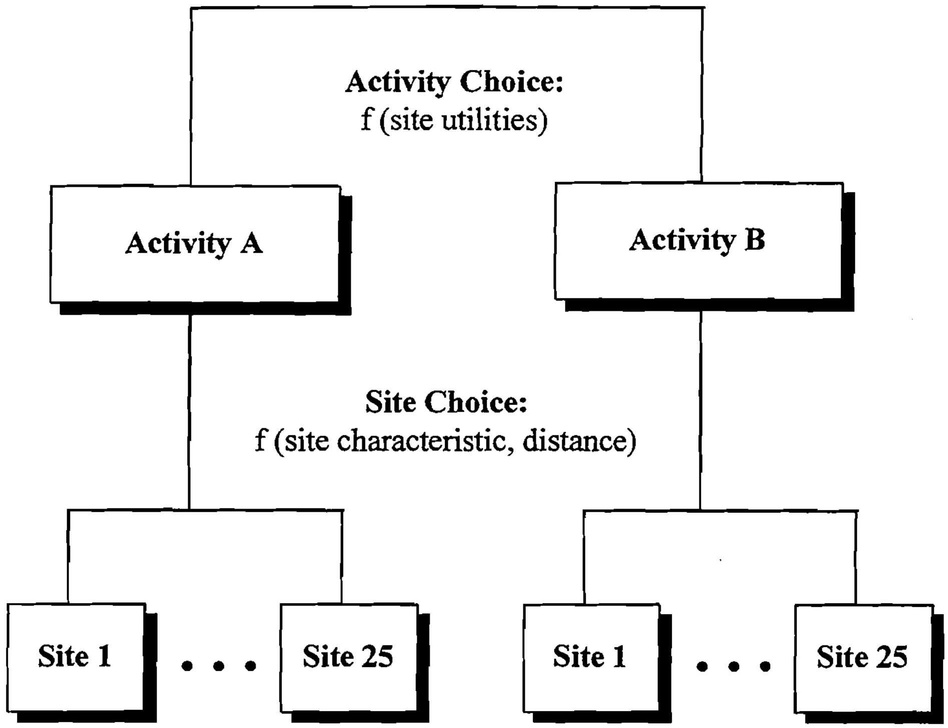


FIGURE 1.
MODEL STRUCTURE

The underlying behavioral model requires specifying the utility that an individual would receive from each alternative in the choice set. Let U_{ni} represent the utility that the individual receives for choosing activity n and site i.[3] U_{ni} is divided into deterministic (V_{ni}) and random (ε_{ni}) components with the deterministic part assumed to be linear in the parameters.

(1) $U_{ni} = V_{ni} + \varepsilon_{ni}$

(2) $V_{ni} = X_{ni}\beta$

where X_{ni} is a vector of site characteristics with associated parameters β. Assuming a generalized extreme value distribution on ε_{ni}, the probability that an individual chooses activity n

[3] For simplicity we have suppressed the notation for individuals.

and site i, P(ni), is the product of the marginal and conditional probabilities presented in Equations 3-5. s_n in Equation 4 follows the notation of Morey (1994) and is equivalent to McFadden's $1/(1-\sigma_n)$.

$$(3) \qquad P(ni) = P(n)P(i|n)$$

$$(4) \qquad P(n) = \frac{a_n \left[\sum_j e^{s_n V_{nj}}\right]^{1/s_n}}{\sum_m a_m \left[\sum_j e^{s_m V_{mj}}\right]^{1/s_m}}$$

$$(5) \qquad P(i|n) = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

Let

x_{ni} = site characteristic value for activity n, site i $\qquad$ n = A,B; i = 1, . . ., 25

d_{ni} = distance to site i for activity n $\qquad$ n = A,B; i = 1, . . ., 25

The model assumes that both activities are available at all 25 sites so $d_{ni} = d_i$.

Site characteristics 1 and 2 are drawn from random uniform distributions with ranges 0 to 250 to 0-1.5, respectively. The two site characteristics are drawn on different uniform distributions in order to simulate differences in characteristics that could affect distinct activities.

The distance variable is randomly assigned in two steps. First the average distance to each site is assigned by drawing on a uniform distribution with range 80 to 160. The second step varies the distance for each trip by adding a normal random number to the average site distance with a standard deviation of 25.[4] Other specific parameters of the model are found in Table 1.

TABLE 1.
MEASUREMENT ERROR AND SAMPLING EXPERIMENTS' PARAMETERS

1000	=	number of trips
25	=	number of sites
2	=	number of activities
0.002	=	site characteristic 1 coefficient
0.16	=	site characteristic 2 coefficient
-0.04	=	distance coefficient
0.7	=	$\sigma_n \quad \forall n$
1	=	$a_n \quad \forall n$

Using the pseudo-data and the parameters presented in Table 1, the "true" P(ni) were generated using Equation 3 for each choice alternative (trip). A sample of trips was then drawn from this distribution for each iteration of the Monte Carlo simulation. The studies presented below are the result of 100 runs of the respective models, each of which is estimated with Full Information Maximum Likelihood.[5]

---

[4] Within these parameters, it is unlikely that a negative distance would ever get assigned to a site for a given trip. If a negative distance was encountered, it was assigned the distance of 2.

[5] Manski (1975) presents Monte Carlo results from 25 independent samples with 400 observations in each sample. This paper presents results from 100 independent samples with 1,000 observations in each sample. Brownstone and Small (1989) base their Monte Carlo results on 100 independent samples showing little empirical change beyond 60 repetitions.

Errors-in-Variables

In an OLS regression model with a single regression, measurement errors in the explanatory variable bias the regressor's estimated coefficient toward zero. When one regressor is measured with error in a multivariate setting, the coefficient of the affected variable again is biased toward zero and the coefficients of the other well-defined variables are also biased but in an indeterminate direction. In Greene's words, "A badly measured variable contaminates all of the least squares estimates."[6] Such is also the case in the RUM.

To investigate the effect of using a variable measured with error, we add random "noise" to the site characteristic for activity A. This noise is a normally distributed random number with a mean of zero and a standard deviation equal to a specified percentage of the overall mean of the site-characteristic distribution. A separate draw from a $N(0, p\overline{X})$ distribution is made for each site in activity A and added to the "true" value, where p is the specified percentage, and $\overline{X}$ is the average site characteristic across all sites. The same site attributes and trip distances are used for each model estimation. However, each iteration draws new noise and trip vectors from the respective distributions. One hundred independent samples are estimated for each of eleven levels of error: 0% through 100% in increments of 10%.

---

[6] Greene 1993, p. 284.

FIGURE 2.
SENSITIVITY OF ESTIMATED SITE CHARACTERISTIC 1
TO MEASUREMENT ERROR

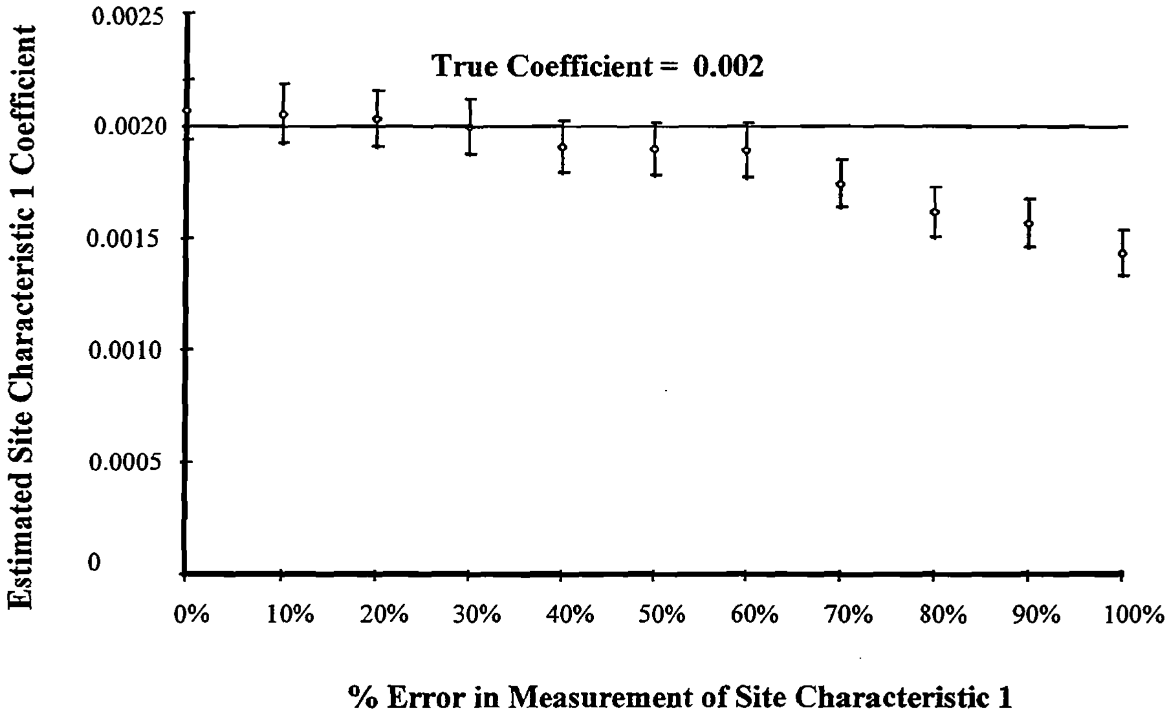
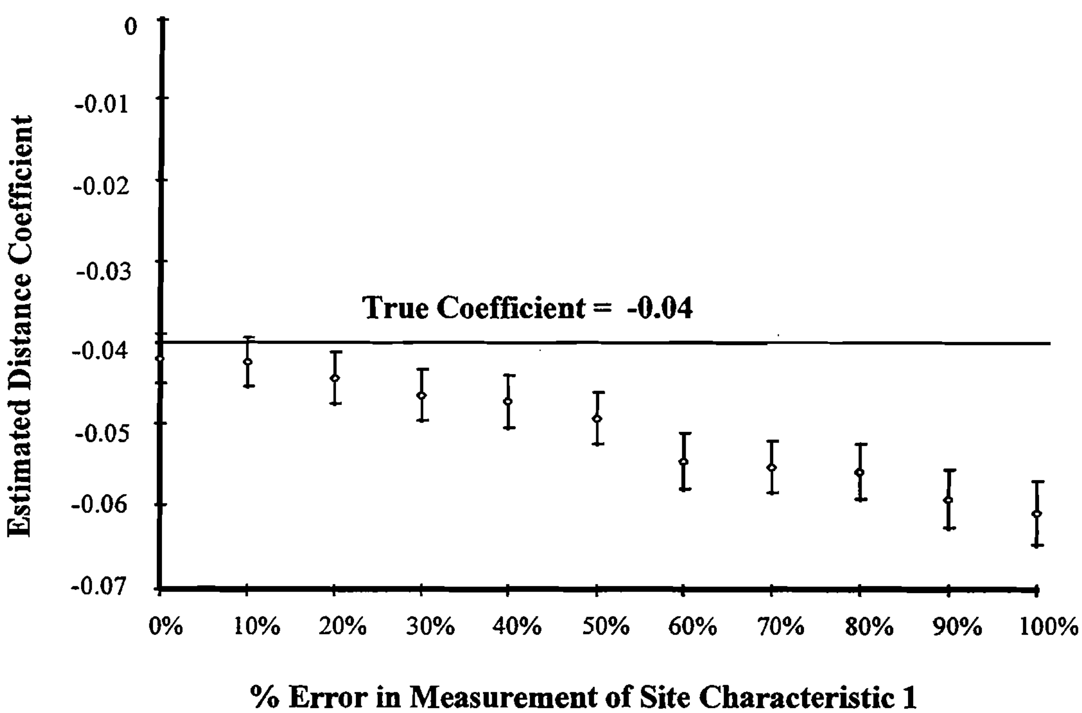


Figure 2 shows the clear attenuation of activity A's site-characteristic coefficient. It is interesting that with 100% error added onto the characteristic, the average estimated site characteristic coefficient across 100 runs of the model is only 28.10% below the true value.

Figure 3 demonstrates the effect of measurement error in site characteristic 1 on the distance coefficient. Clearly, the distance coefficient also moves in an unambiguous direction as a result of the measurement error in site characteristic 1. However, it is noteworthy that the direction of the bias is not toward zero. At 100% error in the site characteristic, the distance coefficient has increased in absolute magnitude almost 53%. Because the site characteristics and distances are independently drawn, the effect on the distance coefficient is not due to correlation between distance and the site characteristic.

FIGURE 3.
INFLUENCE OF SITE CHARACTERISTIC 1
MEASUREMENT ERROR ON THE ESTIMATED DISTANCE COEFFICIENT



The other lower-level coefficient estimated in the model is site characteristic 2 associated with activity B. The average value of this coefficient is within the 90% confidence interval of it's true value when activity A's site characteristic is measured with 100% error. However, the upper-level parameters, s_A and s_B, are affected by the measurement error in site characteristic 1.

These upper-level parameters have true values of 3.33 and are each estimated at around 2.2 at 100% measurement error of site characteristic 1 -- a reduction of approximately 33%.

The random utility model is used in the recreation literature to estimate demand and value of specific resources. Thus the effect of potential parameter biases on willingness-to-pay values is of considerable interest. As shown in Figure 4, the effect of the measurement error in the activity A site characteristic coefficient is an observable reduction in willingness to pay (WTP). This WTP is calculated for a 10% improvement of activity A's site characteristic at all 25 sites. At 0% measurement error, the mean WTP for this improvement is $0.48. The mean WTP continually decreases across the next ten experiments until it is equal to $0.26 at 100% measurement error. This is a 45.8% decrease in WTP.

Sampling

This experiment examines the effects of employing a nonrepresentative sample. In this experiment the five sites with the highest values of site characteristic 1 are oversampled relative to the true distribution of 649 trips. The oversampling occurs by our sampling the 649 trips with replacement until 800 trips were generated from the 5 most popular sites. The 351 trips to the other 20 sites were sampled with replacement until we had 200 trips to those sites. Trips to the oversampled sites were assigned a weight of 649/800, and trips to the undersampled sites were assigned a weight of 351/200. The following four scenarios were then examined:

WW = weighted model estimation, weighted WTP

WU = weighted model estimation, unweighted WTP

UW = unweighted model estimation, weighted WTP

UU = unweighted model estimation, unweighted WTP.

FIGURE 4.
WTP FOR 10% INCREASE IN ACTIVITY A ATTRIBUTE FOR ALL SITES

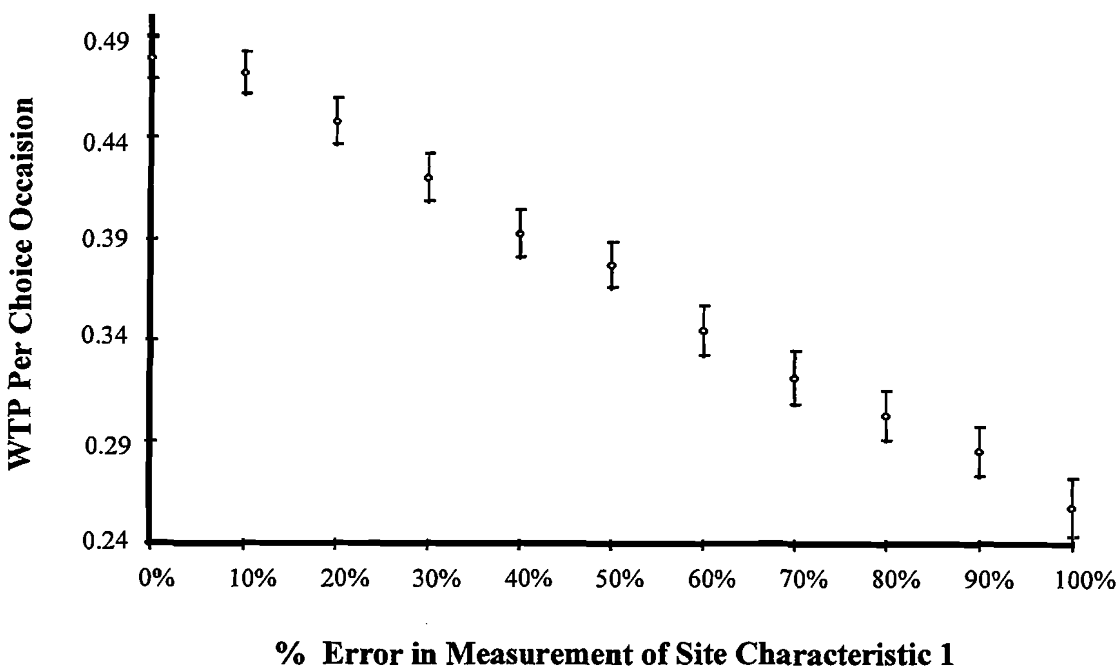
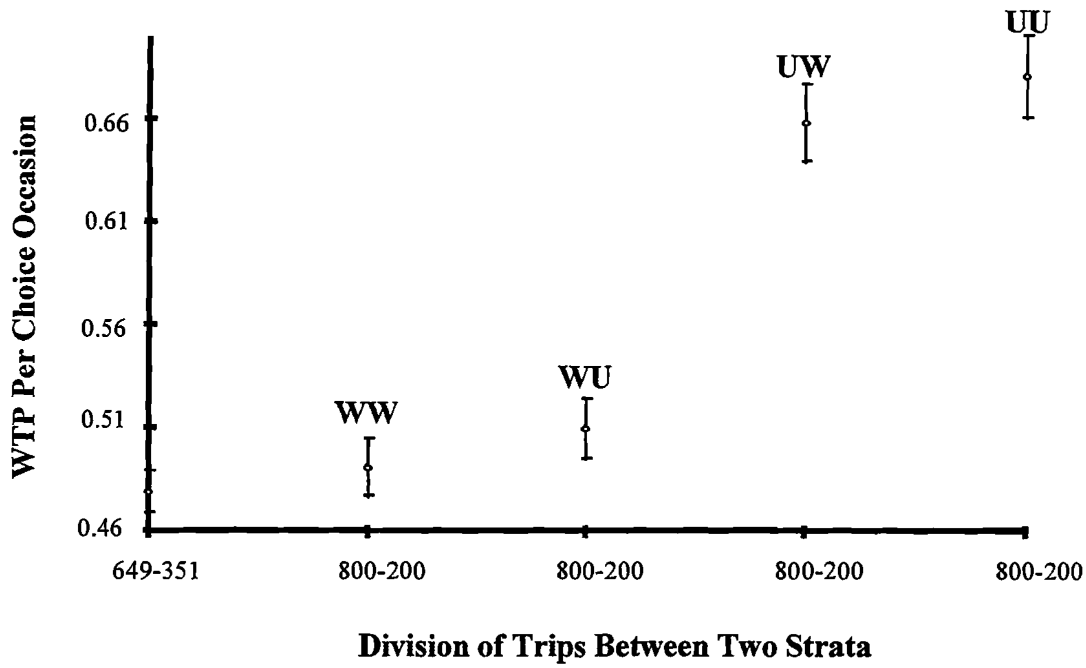


FIGURE 5.
WTP FOR 10% INCREASE IN ACTIVITY A ATTRIBUTE FOR ALL SITES



The WTP calculations are, as in the measurement error discussion above, for a 10% improvement in site characteristic 1 across all 25 sites. Figure 5 presents the WTP calculations across the 100 runs of the model for the four scenarios. As seen in the figure, the biggest effect is in not properly weighting the estimation. Not weighting the WTP calculation has only a second-order effect.

This experiment was repeated by oversampling only the most popular site. The true number of trips to this site was 260 leaving 740 trips to the other 24 sites. The most popular site was oversampled until it accounted for 400 trips in the sample which left a sample of 600 trips to the other 24 sites. In this experiment the standard errors were larger around the WTP values, but outcome showed the same pattern of weighted and unweighted results as displayed in Figure 5.

It is well known that it is necessary to use weights when the sample is not representative of the population. The challenge with unrepresentative samples is the calculation of correct statistical weights. With sample designs similar to this experiment, it is likely that the researcher will know the true trip distribution necessary to calculate the weights. Without statistical weights the researcher must adjust the final mean WTP value ex post. The correct transformation requires information about the nature and magnitude of the average distortion in order to produce an unbiased value estimate. Of course, information of this kind rarely is available.

Limited Site Information

The limited site information experiment differs from the previous experiments in that a two-level nested-logit structure is imposed by the researcher, but is not behaviorally motivated. The true data generating process is a single-level multinomial logit model. However, the model is estimated as a nested-logit model. In the first two experiments, there was a behavioral choice between two activities at the same sites. For this experiment, the activity-level choice is a choice between two categories of sites. In category A the sites are modeled with a single site characteristic and distance. In category B, all that is known is that trips were made to sites for which there is no site information. The nesting structure has the recreator first choose a category of sites. If category A is chosen, then the recreator chooses a site based on its distance and the value of the site characteristic. If category B is chosen, then all that is modeled is that the second category is chosen.

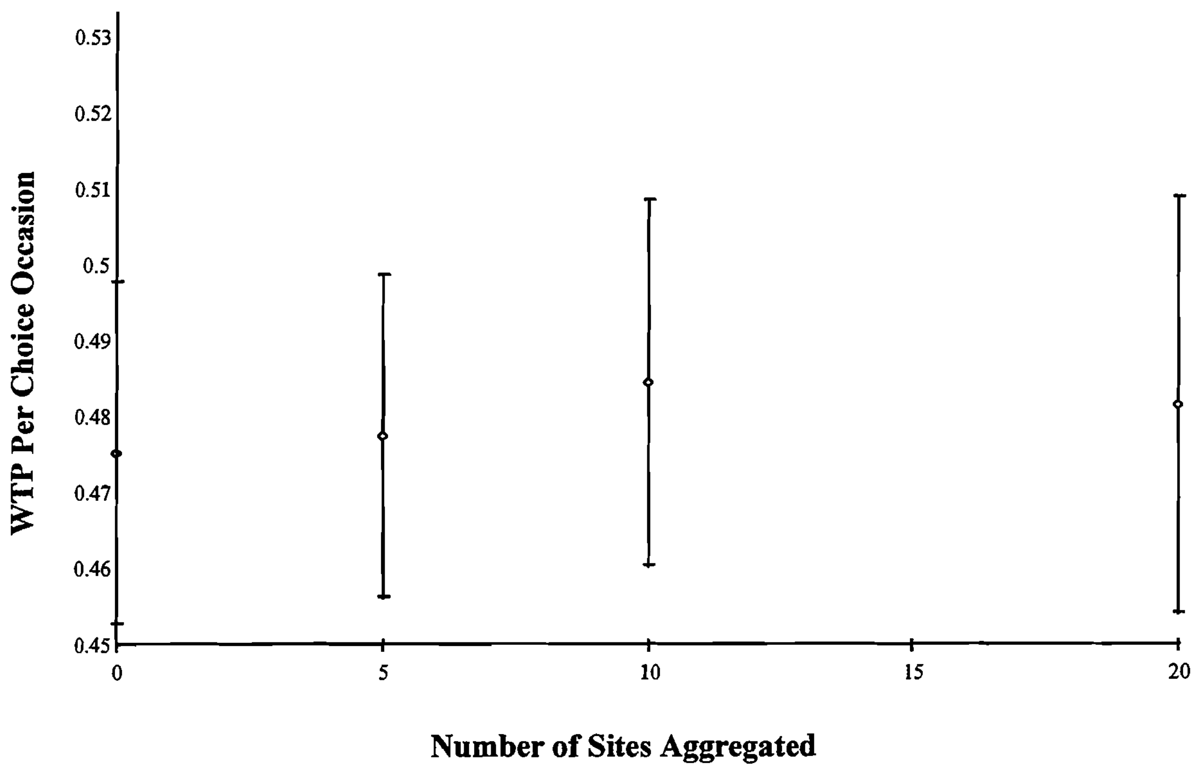
There are 40 sites in this experiment. Each of the 40 sites are assigned a site characteristic and distance which are used to calculate the true trip distribution by the methodology described above. Of the 40 sites, we assign a certain number to be modeled in the second category. In essence, we throw away the site characteristic and distance information on the site and only model the number of trips that are taken in this second category. We might think of this second category as an aggregate site with either undefined or such noncomparable site characteristics that the information is dropped. Category B is estimated with an alternative-specific dummy variable at the category-choice level. The true coefficient for the category B dummy variable is calculated using the true cumulative probabilities for the category B sites.

TABLE 2.
LIMITED SITE INFORMATION PARAMETERS

1000	=	number of trips
40	=	number of sites
0.002	=	site characteristic coefficient
-0.04	=	distance coefficient
σ_n	=	0.7

This experiment assumes that the sites to be included in category B occur randomly. We estimate models with an increasing number of sites in category B. Figure 6 shows the WTP for a 10% increase in a single site's characteristic under the true specification and when 5, 10, and 20 sites out of 40 are included in category B. For each of these levels of aggregation, the results again are based on 100 independent samples.

FIGURE 6.
WTP FOR 100% INCREASE IN ATTRIBUTE FOR SITE 10



The WTP values for each of the two-level nested logit models do not appear to be systematically biased relative to the true specification. Additionally, there has been little loss in efficiency from using the artificially imposed model. This result may be an artifact of having randomly chosen the category B sites. Additional work needs to be done to consider the effects of strategically choosing sites. Possible strategic choices could include putting less attractive sites or more distant sites into the aggregate alternative.

Conclusions

The results presented in this paper generally lend support to OLS intuition. The errors-in-variables experiment confirmed the expected attenuation in the affected site characteristic coefficient and was consistent with Greene's statement that measurement error can contaminate all of the estimates. The sampling experiment demonstrated the well-known result that it is important to use appropriate weights in both the estimation and WTP calculations when the sample is not strictly random. The challenge in such an oversampling strategy is that the researcher often does not know the true trip distribution by which to calculate the proper weights. It also should be apparent that the researcher cannot make a simple adjustment after the unweighted estimation and expect to obtain an unbiased WTP estimate, unless the researcher knows the true WTP from an external source. Finally, the limited site information problem demonstrates the potential for incorporating sites with incomplete information into an aggregate category. Caution needs to be used in generalizing this result given that the aggregated sites were selected randomly.

The unstartling nature of these results should encourage more researchers to explore the rich structure of random utility models in the recreation demand literature. While computationally more challenging, much of the intuition developed in the OLS regression framework will translate over into the nonlinear applications. Our experience has been that when results do go counter to our OLS intuition, we must go back and assure there are no programming or data-entry errors. The payoff of using the RUM to model recreation is seen immediately in the more defensible modeling of substitutes and the decision process.

References

Ben-Akiva, Moshe and Steven R. Lerman. 1985. *Discrete Choice Analysis.* London England: The MIT Press.

Brownstone, David and Kenneth A. Small. "Efficient Estimation of Nested Logit Models." *Journal of Business & Economic Statistics*, 7(1):67-74.

Manski, Charles F. 1975. "Maximum Score Estimation of the Stochastic Utility Model of Choice." *Journal of Econometrics* (3):205-228.

Morey, Edward R. 1994. "TWO RUMs unCloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of participation and Site Choice." Department of Economics, University of Colorado, Boulder.

Morey, Edward R., W. D. Shaw, and R. D. Rowe. 1991. "A Discrete-Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data Are Not Available," *Journal of Environmental Economics and Management*, 20(2): 181-210.

Parsons, G.R. and M.J. Kealy. 1992. "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics,* 68(1):93-106.

Train, Kenneth. 1986. *Qualitative Choice Analysis.* London England: The MIT Press.

Yoshiaki Kaoru, V. Kerry Smith, and Jin Long Liu. 1995. "Using Random Utility Models to Estimate the Recreation Value of Estuarine Resources." *American Journal of Agricultural Economics* (77):141-151.

An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization*

Catherine L. Kling

and

Joseph A. Herriges

March 29, 1995

*Both authors are associate professors of economics at Iowa State University, Ames, Iowa. Partial funding for this research was provided by the Western Regional Project W-133. This is Journal Paper No. J-16090 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa; Project No. 3246. The authors would like to thank Doug Miller, seminar participants at Iowa State University, and two anonymous journal reviewers for constructive comments.

An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization

Nested logit models (NLMs) continue to be popular in empirical studies of recreation site selection and the welfare implications of changes to site amenities (a nonexhaustive list includes Bockstael, Hanemann, and Kling, 1987; Bockstael, McConnell, and Strand, 1989; Jones and Stokes, 1987; Milon, 1988; Morey, Shaw, and Rowe, 1991; Morey, Rowe, and Watson, 1993; Lupi et al., 1994). A primary advantage of this specification is that it yields closed-form equations for the site choice probabilities, thus easing estimation, while allowing for varied correlation patterns among selection alternatives. In addition, nested logit models provide an intuitively appealing structure for the sequence of decisions involved in site selection, much like the two-stage budgeting process employed in consumer demand theory. They also deal well with issues of non-participants, choices among different types of recreational sites and activities, and different types of users.[1] For these same reasons, the nested logit structure is also widely used in other areas of applied economics, most notably in transportation mode and housing demand analyses.

One issue in the empirical implementation of these models is whether the estimated models are consistent with utility maximization. In particular, it has been shown that the inclusive value coefficient (or dissimilarity coefficient) must lie in the unit interval for the empirical model to be globally consistent with utility maximization (Daly and Zachary, 1979; McFadden, 1981). Unfortunately, it is not uncommon to observe estimates of these coefficients that exceed one. In this situation, researchers have generally followed one of two alternatives. They might interpret these results as a rejection of the model specification and re-estimate the model with a different set of explanatory variables or nesting structure. Alternatively, they might proceed with the model interpreting the coefficients in a purely statistical sense where the size of the dissimilarity coefficient simply represents the degree of substitutability between the alternatives (see, e.g., Train, Ben-Akiva, and Atherton, 1989). In this case, no concern is directed towards whether the empirical model is consistent with utility maximization.

Recently, Börsch-Supan (1990) has provided a third alternative. Therein, he argues that the global conditions of Daly and Zachary (1979) and McFadden (1981) (hereafter referred to as the DZM conditions) are too stringent. Specific nested logit models, like flexible functional forms used in demand analysis, should be viewed as providing a local approximation to the true underlying model. As a result, conditions for consistency with utility maximization should only be expected to hold locally, such as at the mean of the sample or over a relevant range of choice probabilities. Börsch-Supan then derives a set of conditions under which a nested logit model is consistent with utility maximization, with dissimilarity coefficients that lie outside the unit interval. In an extension and correction to Börsch-Supan's work, Herriges and Kling (1994) identify additional restrictions when nesting structures contain more alternatives than Börsch-Supan considers. However, neither Börsch-Supan nor Herriges and Kling discuss the empirical implementation of these conditions. The purpose of this paper is to present several implementation strategies and to provide an example of the methods using a recreational sportfishing application.

The paper is organized into five sections. In the first section, we present the basic nested logit model and the necessary conditions for its global and local consistency with utility maximization. The data and empirical model employed in the example are then described in the second section. In the third section, we present three alternative approaches to checking for the consistency conditions, providing tests of those conditions based on classical statistics. A Bayesian approach to interpreting and imposing the local consistency conditions is provided in the fourth section. Final remarks and conclusions follow in the last section.

Nested Logit Models and the Consistency Conditions

The development of the NLM was largely in response to concern about the independence of irrelevant alternatives (IIA) assumption imbedded in traditional multinomial logit models. This assumption requires that the relative choice probabilities for any two alternatives be independent of the other choices available (See McFadden, 1981, pp. 221-22). However, tests of the IIA assumption in the empirical literature frequently reject this limitation on preferences (e.g., Börsch-Supan, 1987, Train, McFadden, and Ben-Akiva, 1987). The NLM specification relaxes the IIA assumption of traditional logit models by introducing one or more dissimilarity coefficients (θ_k's). These coefficients allow for varying patterns of substitution among alternatives in the choice set. Researchers may hypothesize that certain groups of alternatives are more similar than others and combine these alternatives together in one nest. For example, in a recreational fishing context, researchers might group alternative fishing sites together by choice of mode (i.e., fishing from the beach, pier, private boat, or party boat).

Formally, the probability (P_i) that an individual will select alternative i is divided into two components in the NLM, with

$$(1) \qquad P_i(v) = Q_{m(i)} P_{i|m(i)}$$

where v_i denotes the utility associated with alternative i, $v \equiv (v_1, \ldots, v_N)$, m(i) maps the alternative (i) into the mode (or nest) to which the alternative has been assigned by the analyst, and M denotes the total number of modes. Q_k denotes the probability that mode k is selected and $P_{i|m(i)}$ represents the conditional probability that alternative i is selected given that mode m(i) has been chosen, with

$$(2) \qquad P_{i|m(i)} = \frac{exp(v_i / \theta_{m(i)})}{E[m(i)]}$$

and

$$(3) \qquad Q_{m(i)} = \frac{E[m(i)]^{\theta_{m(i)}}}{\sum_{k=1}^{M} E[k]^{\theta_k}}$$

where

$$(4) \qquad E[k] = \sum_{\{i|m(i)=k\}} exp(v_i / \theta_k) \ .$$

θ_k is the dissimilarity parameter (or inclusive value coefficient) associated with mode k. If $\theta_k = 1, \ \forall \ k = 1, \ldots, M$, then the straight multinomial logit model results and the IIA assumption holds for all alternatives. Often these models are estimated imposing the assumption that $\theta_k = \theta,\ \forall\ k$ (Morey, 1994), but this restriction is not necessary and is easily relaxed.

As previously noted, DZM identify the set of conditions on parameter values under which the NLM (1) is globally consistent with utility maximization. The most problematic restriction from an empirical perspective is that the choice probabilities, P_i, must have nonnegative even and nonpositive odd mixed partial derivatives with respect to components of v other than v_i. This condition ensures that the distribution has a nonnegative density.[2] In order for this condition to hold globally, it can be shown that the dissimilarity coefficients must lie inside the unit interval (McFadden, 1979; Daly and Zachary, 1979); i.e.,

(5) $\quad 0 < \theta_k \leq 1,\ \forall k.$

The key contribution of Börsch-Supan was to note that, if consistency with utility maximization is required to hold only locally, then the θ_k's can lie outside of the unit interval. Specifically, it can be shown (Börsch-Supan, Theorem 1, and Herriges and Kling, Corollary 1) that the nonnegativity constraint on the first order partial derivatives of P_i implies the following necessary condition for all nests with two or more alternatives:

(6) $$\theta_k \leq U_{1k}(v) \equiv \frac{1}{1 - Q_k(v)} \qquad k = 1,\ldots,M.$$

Note that since $0 \leq Q_k \leq 1$, this condition is always satisfied for any value of $\theta_k \in [0,1]$. However, for the condition to be satisfied when $\theta_k > 1$, Q_k must be sufficiently large. This necessary condition can be easily checked as there is only one condition for each nest and since one need only to compute the Q_k's and θ_k's, both of which are easily recoverable from model output.

Similarly, the nonpositivity restriction on the mixed second partial derivatives of P_i are shown in Herriges and Kling to impose the following condition on θ_k's for nests with three or more alternatives:

$$(7) \qquad \theta_k \leq U_{2k}(v) \equiv \frac{4}{3[1-Q_k(v)]+\sqrt{[1+7Q_k(v)][1-Q_k(v)]}} \ .$$

Again, there is again only one condition to check for each nest.

In general, for a nest with n alternatives, there will be n-1 necessary conditions for each order of mixed partial derivatives. For example, with six alternatives in a nest, there will be five sets of conditions to check for that nest. Likewise, if there are only two alternatives, there is only a single condition to check corresponding to (6), so that $\theta_k \leq U_{1k}(v)$ is both necessary and sufficient. In practice, even when a model has many alternatives within each nest, it may only be practical or desirable to check the first and possibly the second derivatives. That is, given the errors implicit in model estimation, satisfaction of the necessary condition (6) may be considered adequate. This is analogous to the approach often used in demand analysis of checking only the sign of the own price elasticities or the diagonal elements of the matrix of Allen elasticities of substitution.

Before proceeding with the discussion of the empirical implementation, it is worth commenting on the merits of requiring consistency with utility theory when performing welfare analysis. There are two strands of thought in this regard in the environmental valuation literature. One approach is to argue that empirical welfare measurement is meaningful only if the resulting welfare esimates are consistent with the postulates of the welfare theory underlying them. Examples of papers and the requirements they address include: Hicksian versus Marshallian welfare measures (Bockstael and McConnell, 1993), the satisfaction of the integrability conditions (Bockstael, Hanemann, and Strand, 1986), the constraint that willingness to pay measures not exceed income (Kling and Sexton, 1992, and Kling, 1991), and consistency with the theoretical requirements of demand systems (Hanemann and Morey, 1992).

An alternative view argues that empirical welfare measures are by their nature approximations. This view suggests that there is so much error implicit in estimation that welfare measures that are not entirely consistent with utility theory are fine. Further, this view suggests that the best strategy is to find the model that best fits the data and use it to compute welfare measurement, regardless of whether the resulting estimates are consistent with utility theory.

In the following sections, we describe the data, empirical specification, and approaches to testing consistency with utility theory. The degree to which the researcher would want to use these approaches as tools for model selection depends on where one comes out in the previous debate. In our view, welfare evaluation using recreation demand continues to be part art and part science. Examination of whether estimated models are consistent with utility theory can be combined with traditional goodness-of-fit tests to help evaluate models. Researchers should report the results of such evaluations, but must still use a dose of good judgement in the final analysis.

Data and Empirical Specification

The model estimated and data employed here are more fully described in Thomson and Crooke (1989) and Kling and Thomson (1994). In Kling and Thomson, a variety of NLMs are estimated that differ in their nesting structure and the implications of the various structures for welfare measurement are investigated. Here, one of the nesting structures implemented and estimated in that study is used as a model for applying and testing the consistency conditions.

The data used to estimate the model are from the Southern California Sportfishing Recreation Survey conducted in 1989 (Thomson and Crooke, 1991). In this survey, the general population of eight coastal counties of Southern California was randomly sampled to identify anglers who were then mailed a follow-up survey. Mail survey respondents provided personal information on their income, zip code of residence, and other socioeconomic characteristics. Extensive information regarding their most recent saltwater fishing trip was solicited including the site they visited, mode of fishing (pier, beach, private boat, or charter boat), target species, travel distances, travel time, and expenditures. Roundtrip travel costs to each site were computed by multiplying roundtrip distances by a constant cost per mile and adding an opportunity cost of travel time, costs associated with chartering a boat, and fuel expenses where appropriate.

Data on catch rates were obtained from the Marine Recreational Fishery Statistics Survey (MRFSS) which provided catch rates on a per hour fished basis for each major species group by fishing mode and fishing area. Mail survey respondents frequently identified several target species for their most recent saltwater experience. Target species included bass, barracuda, bonito, rockfish, lingcod, shark, flatfish, and other. A catch rate variable was constructed by summing the per hour catch rates associated with each angler's target species. Since the MRFSS data was collected independently from the trip data, the catch rate associated with each alternative is exogenous to the trip decision.

Upon combining the MRFSS creel data with the angler survey, there were a total of 26 mode/site alternatives: five sites that can be fished from the beach, five pier sites, and eight sites that could be accessed by private boat or charter boat. While there are a myriad of possible nesting structures one could consider, the nesting structure used in the present study groups alternatives by mode of fishing, as illustrated in Figure 1. This structure is chosen for two principle reasons: (1) it is a commonly estimated form for sportfishing studies and (2) it yields coefficients for the dissimilarity coefficients that lie outside the unit interval. The indirect utility associated with each alternative was assumed to be a simple linear function of catch rates and travel cost, with

(8) $$v_i = \alpha + \beta_p T_i + \beta_c C_i$$

where T_i denotes the roundtrip travel cost associated with site i and C_i denotes the catch rate at site i.

The nesting structure in Figure 1, together with equations (1) through (4) and (8), can then be used to formulate the appropriate log likelihood function and estimate the parameters of the model. In this paper, TSP's ML procedure was employed. The resulting coefficient estimates are provided in the second column of Table 1.

Both the price and catch rate coefficients have the expected signs and are significantly different from zero at a one percent significance level. Two of the four dissimilarity coefficients lie in the unit interval. However, the coefficient for private and charter boat models are 1.19 and 1.93, respectively. Thus, global consistency with utility maximization, as defined by the DZM condition (5), is violated for these two cases. A simple extension of the DZM condition is to recognize that the $\hat{\theta}_k$'s are random variables and to test whether they are significantly greater than 1. Doing so generates t-ratios of 1.84 and 6.86, respectively, for the private and charter boat modes, also suggesting rejection of the DZM conditions for these two modes. Using the asymptotic normality of the maximum likelihood parameter estimates, the probability that the θ_k's jointly lie within the unit interval is less than .1%. These results suggest that the model as estimated is not globally consistent with utility maximization. In the following section, we apply the Börsch-Supan conditions to check whether, despite the global inconsistency, the estimates are consistent locally with utility maximization.

Classical Tests of Consistency with Utility Maximization

The crux of Börsch-Supan's original argument was that consistency with utility maximization should be required to hold only for "...data points that are sensible for a specific application of the choice model." (1990, p. 377). However, the selection of the domain over which the conditions will be required to hold is left as a task for the analyst. In this section, we identify three classical approaches to implementing and testing the consistency conditions identified in (6) and (7). Briefly, the first two methods require the consistency conditions to hold : (1) for all observations in the sample (or some proportion of observations) versus (2) at the means of the explanatory variables. The third approach recognizes that, even if the point estimates fail to satisfy the local consistency condition, there may not be sufficient precision in the parameter estimates to statistically reject consistency. Thus, we propose testing the conditions using the mean of the sample's explanatory variables.

The first approach takes one extreme in checking whether the model satisfies the consistency conditions, namely whether the first and second order conditions hold for each of the 1182 observations in the data set. The second column of Table 2a reports the percentage of the time that the first order condition is satisfied in the sample for each fishing mode. Thus, for all but the charter mode, the first order condition that $\theta_i \leq U_{1i}(v)$ is satisfied for the entire sample. Clearly this result is expected for the beach and pier modes, which satisfy the global DZM condition of $\theta_i \leq 1$. The private mode result is encouraging in that, despite the violation of the DZM restriction with $\theta_{private} = 1.19$, we still find 100 percent compliance with the first order condition. The charter mode results, on the other hand, are not as encouraging with the first order condition being satisfied for less than eight percent of the sample. A similar story emerges when we turn to the second order conditions, as indicated in Table 2b. While the second order conditions are uniformly satisfied for both beach and pier modes, they are rarely satisfied for either the private or charter modes. The higher probabilities associated with choosing these modes are simply not large enough to counter the large estimates of the dissimilarity parameters.

The first approach may be viewed as too stringent in that, while global consistency is not sought, consistency is expected to hold for the entire domain of the explanatory variables. This is rarely expected of continuous demand models when flexible functional forms are employed. Instead, analysts typically check curvature conditions only at the mean of the sample. This leads to our second approach to applying the local consistency conditions, checking whether or not the dissimilarity coefficients satisfy $\hat{\theta}_k \leq \hat{U}_{1k}(\bar{v})$ in the case of the first order conditions and $\hat{\theta}_k \leq \hat{U}_{2k}(\bar{v})$ in the case of the second order conditions, where $\bar{v}$ indicates the value of the indirect utility function evaluated at the means of the explanatory variables.

Values for $\hat{U}_{1k}(\bar{v})$ and $\hat{U}_{2k}(\bar{v})$ are reported in the fourth column of Table 2, with the $\hat{\theta}_k$'s repeated in column 3 for comparison sake. Qualitatively, the implications for local consistency with utility maximization do not change from the first approach. The first and second order conditions hold for the beach and pier modes. For private, the first order condition holds, but not the second and for charter neither condition holds. Based on the results of these first two evaluation methods, a clear case cannot be made for whether the empirical model is consistent with utility maximization. However, as one would expect, the second method's results are closer to suggesting compliance. Given that the parameter estimates are just that, estimates, the question is whether or not there is sufficient precision in the coefficients to reject local consistency. This leads to our third and final method for checking the first order local consistency conditions, testing the hypothesis:

$$H_{10}: \theta_k \leq U_{1k}(\bar{v})$$

against the alternative hypothesis

$$H_{1A}: \theta_k > U_{1k}(\bar{v})$$

for each k. Note that these are one-tailed tests. Likewise, hypotheses associated with the second order consistency conditions can be written:

$$H_{20}: \theta_k \leq U_{2k}(\bar{v})$$
$$H_{2A}: \theta_k > U_{2k}(\bar{v}) .$$

Results of these tests are reported in the last column of Table 2, where the t-ratios associated with the test statistic

(9) $\hat{z}_{jk} \equiv \hat{\theta}_k - \hat{U}_{jk}(\bar{v})$

are reported. In both cases, negative t's immediately signify failure to reject the null hypothesis (i.e., compatibility with utility maximization). It is only t's that indicate significant differences from zero in the positive direction that suggest rejection. Based on this criteria, all four of the modes satisfy the first order condition, but for only one of the four modes (charter) do we reject the second consistency conditions at any reasonable confidence level. These results are more conclusive, particularly if satisfaction of the first order condition is felt to be adequate.

While individual tests of the consistency conditions is revealing, what would be preferred is a joint test for all modes; i.e., a test that all of the first derivative conditions taken together are satisfied and a test that all of the first and second derivative conditions are jointly satisfied. Since the methods used parallel those for the Bayesian approach, these joint tests and their results are discussed in the next section.

A Bayesian Approach to Imposing Consistency with Utility Maximization

In the previous section, our focus was on an ex-post evaluation of the model to determine whether it is consistent with the first and second order necessary conditions for consistency with utility maximization. An alternative approach is to impose these conditions in estimating the coefficient vector. To do so, we adopt the Bayesian perspective of combining prior beliefs about the distribution of the coefficient vector with evidence from the sample. In this case, our priors are summarized by the necessary conditions, either condition (6) or conditions (6) and (7) simultaneously. For example, for consistency with condition (6), we adopt the uniform prior density function

$$(10) \qquad f_1(\beta,\theta) \propto \begin{cases} 1 & \theta_k \le U_{1k}(\bar{v}) \quad \forall k \\ 0 & otherwise \end{cases}.$$

Following Bayes rule for continuous functions, the posterior density function can then be expressed as

$$(11) \qquad f_1(\beta,\theta|X) \propto f_1(\beta,\theta)L(\beta,\theta|X),$$

where X is the set of explanatory variables and $L(\beta,\theta|X)$ is the joint density or likelihood function for the sample.

Likewise, priors requiring that conditions (6) and (7) both be satisfied can be summarized as[3]

(12) $$f_2(\beta,\theta) \propto \begin{cases} 1 & \theta_k \leq U_{2k}(\bar{v}) \quad \forall k \\ 0 & otherwise \end{cases}.$$

To combine our priors with the sample likelihood, we employ Monte Carlo integration methods (Kloek and van Dijk, 1978; Geweke, 1986). In this procedure, a large number of draws are taken from the posterior distribution and the mean of these draws is used to estimate the posterior means of parameters. Thus, the posterior density is defined only over the parameter space consistent with the restrictions.

Procedurally, the approach is as follows:

1. Estimate the nested logit model without prior information yielding a parameter vector that is distributed multivariate normal with mean $\hat{\varphi} \equiv (\hat{\beta},\hat{\theta})$ and variance-covariance matrix $\hat{\Omega}$.
2. Randomly resample a large number of times from the $N(\hat{\varphi},\hat{\Omega})$ distribution.
3. Retain only those draws from this distribution that satisfy the priors. This truncated sample provides a series of random draws from the posterior distributions, $f_i(\beta,\theta|X)$, which can in turn be used to characterize the posterior distribution. In particular, the truncated sample can be used to form posterior means for θ_k, ones

that are consistent with the priors. The truncated sample can also be used to estimate probabilities and standard deviations associated with the posterior density function. In the current application, we form the posterior distribution by resampling one million times from the maximum likelihood coefficient vector. In each repetition, consistency with utility maximization is determined by evaluating the conditions (6) and/or (7) at the mean of the explanatory variables.

We begin by considering priors based only on the first order conditions (i.e., $f_1(\beta,\theta)$). Out of the one million draws from the appropriate normal distribution, roughly 13.5% satisfied all four of the first conditions.[4] The means for the resulting posterior distribution, $f_1(\beta,\theta|X)$ are given in column three of Table 1, along with standard errors for each parameter. Notice that all of the parameter estimates are affected by the imposition of the first order priors. As one might expect, both of the θ_k's associated with private and charter modes shrink relative to the classical estimates, though neither falls below one. The estimated θ_k's for both the pier and beach modes are likewise reduced. In addition, the price coefficient becomes less negative and the catch rate coefficient falls slightly. In all cases, the standard error associated with a parameter is smaller as a result of truncating the parameter distribution.

The shifts in the parameter coefficients are more dramatic when the second order condition priors are imposed. From the initial one million draws, only nine (.0009%) of them also satisfy the second order condition in equation (7). The corresponding means for the posterior parameter distribution are provided in column four of Table 1. In this case, only the θ_k associated with the charter mode remains above 1. Furthermore, the price coefficient has shrunk to almost half its level relative to the classical estimates.

We had also hoped to report point estimates that impose the priors of the DZM conditions, i.e., all of the dissimilarity coefficients being contained in the unit interval. However, none of the repetitions we performed provided coefficients consistent with this restriction, thus we estimate a near zero probability of the DZM conditions holding.

In estimating nested logit models of recreation behavior, analysts are often interested in constructing welfare measures associated with the resource. Thus, it is of interest to determine how much welfare estimates change as a result of imposing priors on the coefficients. To examine this, we have constructed welfare measures associated with closing groups of the sites using the standard formula for compensating variation in nested logit models (Hanemann, 1982). Welfare estimates associated with three alternative priors are reported in Table 3. Using the coefficients obtained from a diffuse prior (i.e., no restriction on the coefficients), the average welfare loss associated with closure of all of the shore sites is \$8.40 per choice occasion. (Note that this is also the typically reported classical welfare measure). The average welfare loss associated with closure of all of the offshore sites is \$27.46 per choice occasion.

Using the coefficients when the first order condition priors have been imposed (i.e., column three of Table 1), we obtain shore and offshore estimates of \$10.28 and \$31.57, respectively. Finally, when using the point estimates with priors of both the first and second conditions, we find estimates of \$44.45 and \$16.55 for the offshore and shore sites. The welfare valuations when the local utility maximization priors are imposed have potentially serious implications for policy makers in this case, nearly doubling the shore valuation and increasing the offshore valuation by over sixty percent. The welfare measures rise in these cases primarily due to the reduction in the absolute value of the price coefficient in each case.

Given the sizable differences in welfare estimates associated with imposing the priors, it is reasonable to ask which of these point estimates provide the best guidance for policy makers. Unfortunately, there is no easy answer since, by their nature, Bayesian estimates depend critically on the assumed priors and there are a wide range of priors that might be reasonably imposed. It is our sense that imposition of the first order conditions evaluated at the mean is a sensible minimum restriction and mimics the requirements imposed in other areas of applied demand analysis. However, reasonable arguments can be made in favor of more or less restrictive priors. Thus, we suggest the best course of action is to consider a range of priors such as we have done here (diffuse priors, first order condition priors, first and second order condition priors, etc.), so that individual readers and policy makes have available the information to make their own assessment.

Also, in evaluating the welfare measures, it is useful to contrast these estimates to those found elsewhere for similar resources. For example, in a study estimating the compensating variation associated with closure of the Atlantic Salmon fishery at the Penobscot River in Maine, Morey, Rowe, and Watson (1993) found estimates ranging from about $75 to $150 per trip. Morey, Shaw, and Rowe (1991) found annual compensating variations lost from closing shore modes and boat modes along the Pacific Coast to range from about \$6 to \$110 and from about \$3 to \$60, respectively, depending upon the county of origin. As a final comparison, Bockstael, McConnell, and Strand estimate per trip access values of between \$0.80 and \$8.00 for sportsfishing off the Florida Coast.

Note that the above procedure can also be used to construct the classical joint tests mentioned above of the simultaneous satisfaction of the four first derivative conditions and the joint satisfaction of the eight first and second derivative conditions. In particular, the simulation results reported above can be interpreted in a classical setting as indicating that over a large number of samples, the first derivative conditions will be satisfied roughly 13.5 percent of the time. With standard significance levels of 1 or 5%, this result suggests acceptance of the null hypothesis of consistency with the first derivative condition. In contrast, the results combining the first and second conditions suggest that if a large number of samples were drawn, all eight of the conditions would be satisfied in only about .0009% of them, suggesting rejection of the null that all eight of the first and second conditions are satisfied. These results are quite consistent with the individual tests of the conditions as the t's on the second condition for private and charter modes are quite large, suggesting clear rejection of satisfaction of the second necessary condition.

Final Remarks

Researchers employing NLMs in recreation demand models and other applications have often produced estimates of dissimilarity coefficients that lie out of the DZM bounds for global consistency with utility theory. This study has demonstrated that it is possible to take advantage of the conditions developed by Börsch-Supan and to determine whether the dissimilarity coefficients estimated are locally consistent with utility theory. Based on formal tests of these conditions, this paper has found that although neither the DZM conditions, nor the Börsch-Supan conditions hold for all the modes for the first condition, when framed as a statistical test, the null hypothesis of consistency with the first condition cannot be rejected. Thus, the Börsch-Supan conditions do provide some help in extending the range of the dissimilarity coefficient values that can be interpreted as being consistent with utility maximization.

In this application, we have chosen to concentrate on only the first two derivative conditions. As noted, there will in general be one less derivative condition for each nest than the number of alternatives in that nest. Thus, testing only the first two conditions can be viewed as necessary, but not sufficient. It is our sense that in most circumstances, researchers will be quite content with this level of consistency. In fact, in many cases satisfaction of the first derivative condition may be deemed to be satisfactory.

A final point to note is that consistency with utility maximization is only one possible selection criteria for choosing a model. Consistency with utility theory alone would not likely suffice as a reason for choosing a particular nested structure over another. Traditional model selection criteria such as goodness-of-fit and reasonableness of the chosen variables should continue to be key in choosing model structure.

Table 1: Classical and Bayesian Point Estimates
Under Three Alternative Priors

		Priors Based On:	
Parameter	Classical or Diffuse Priors[1]	First Order Conditions	First and Second Order Conditions
β_p	-0.045 (0.004)	-0.038 (0.002)	-0.025 (0.001)
β_C	0.632 (0.095)	0.591 (0.093)	0.530 (0.094)
$\theta_{private}$	1.19 (0.11)	1.03 (0.05)	0.74 (0.04)
$\theta_{charter}$	1.93 (0.17)	1.68 (0.09)	1.18 (0.05)
θ_{pier}	0.98 (0.14)	0.79 (0.08)	0.45 (0.06)
θ_{beach}	0.88 (0.11)	0.72 (0.06)	0.42 (0.07)

[1] Standard errors are reported in parentheses below the point estimates.

Table 2: Tests of the Consistency with Utility Maximization

(a) First Order Conditions

Mode	Sample % with $\theta_k \leq U_{1k}(v)$	$\hat{\theta}_k$	$\hat{U}_{1k}(\bar{v})$	t-ratio
Private	100%	1.19	1.48	-2.69
Charter	7.6%	1.93	1.76	1.01
Pier	100%	0.88	1.13	-2.38
Beach	100%	0.98	1.14	-1.19

(a) Second Order Conditions

Mode	Sample % with $\theta_k \leq U_{2k}(v)$	$\hat{\theta}_k$	$\hat{U}_{2k}(\bar{v})$	t-ratio
Private	4.5%	1.19	1.14	0.53
Charter	0.0%	1.93	1.24	4.08
Pier	100%	0.88	1.02	-1.33
Beach	100%	0.98	1.02	-0.31

Table 3: Welfare Estimates Associated with Alternative Priors

	Welfare Measure With		
Sites Closed	Diffuse Priors	First Condition as Priors	First and Second Conditions as Priors
Offshore	$27.46	$31.57	$44.45
Shore	$ 8.40	$10.28	$16.55

REFERENCES

Bockstael, N.E., W.M. Hanemann, and C.L. Kling. "Measuring Recreational Demand in a Multiple Site Framework," *Water Resources Research* 23(1987): 951-960.

Bockstael, N. E., W. E. Hanemann, and I. E. Strand, "Measuring the Benefits of Water Quality Improvements Using Recreational Demand Models," EPA Report under Cooperative Agreement #CR-81143-01-1, Washington D.C., 1986.

Bockstael, N.E., and K.E. McConnell, "Public Goods as Characteristics of Non-Market Commodities," *Economic Journal*, (1993).

Bockstael, N.E., K.E. McConnell, and I.E. Strand. "A Random Utility Model for Sportfishing: Some Preliminary Results for Florida," *Marine Resource Economics* 6(1989): 245-260.

Börsch-Supan, Axel. *Econometric Analysis of Discrete Choice With Applications on the Demand for Housing in the U. S. and West-Germany* (Berlin: Springer-Verlag, 1987).

Börsch-Supan, Axel. "On the Compatibility of Nested Logit Models with Utility Maximization." *Journal of Econometrics* 43(1990): 373-388.

Daly, A. and S. Zachary. "Improved Multiple Choice Models," in D. Hensher and Q. Dalvi, eds. *Identifying and Measuring the Determinants of Mode Choice*, Teakfield, London, (1979): 335-357.

Geweke, J. "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics* 1(1986): 127-141.

Hanemann, W.M. "Applied Welfare Analysis with Qualitative Response Models," California Agricultural Experiment Station, 1982.

Hanemann, W. M., and E. R. Morey, "Separability, Partial Demand Systems and Consumer Surplus," *Journal of Environmental Economics and Management*, 22(1992): 241-58.

Herriges, J.A. and C.L. Kling. "Testing the Consistency of Nested Logit Models with Utility Maximization," Staff Papers Series #259, Department of Economics, Iowa State University, 1994.

Jones and Stokes Associates, "Southcentral Alaska Sport Fishing Economic Study," Prepared for Alaska Department of Fish and Game, November 1987.

Kling, C. L., "Estimating the Precision of Welfare Measures," *Journal of Environmental Economics and Management,* 21(1991):244-259.

Kling, C. L., and R.J. Sexton, "Bootstrapping in Applied Welfare Analysis," *American Journal of Agricultural Economics*, 72(1992):406-418.

Kling, C.L. and C.J. Thomson, "The Implications of Model Specification for Welfare Estimation in Nested Logit Models," Draft manuscript, Iowa State University, (1994).

Kloek, T. and J.K. van Dijk. "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica* 46 (1978):1-19.

Lupi, F., H.Z. Chen, T. Tomasi, J.P. Hoehn. "Site-Specific Qualitative Variables in Random utility Travel Cost Models," Paper presented at the American Agricultural Economics Association meetings, (1994).

McFadden, D. "Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments," in: David Hensher and P. Stopher, eds., *Behavioral Travel Modeling* (Croom Helm, London, 1979), pp. 279-318.

McFadden, D. "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications* (1981): 198-272.

Milon, J.W. "A Nested Demand Shares Model of Artificial Marine Habitat Choice by Sport Anglers," *Marine Resource Economics* 5(1988): 191-213.

Morey, E. "Two RUMs UnCLOAKED: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice," Discussion paper, Department of Economics, University of Colorado, Boulder, February 1994

Morey, E.R., R.D. Rowe, and M. Watson. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing," *American Journal of Agricultural Economics* 75(1993): 578-592.

Morey, E. R., D. W. Shaw, and R. D. Rowe, "A Discrete Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data are Not Available," *Journal of Environmental Economics and Management*, 20(1991): 181-201

Thomson, C.J. and S.T. Crooke. "Results of the Southern California Sportfish Economic Survey," NOAA Technical Memorandum, National Marine Fisheries Service, Southwest Fisheries Science Center, August, 1991.

Train, K.E., Ben-Akiva, M. and T. Atherton. "Consumption Patterns and Self-Selecting Tariffs," *The Review of Economics and Statistics* 71(1989): 62-73.

Train, K.E., D.L. McFadden, and M. Ben-Akiva. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *Rand Journal of Economics* 18 (1987): 109-123.

FOOTNOTES

[1] We would like to thank an anonymous referee for pointing out these advantages of nested logit models.

[2] The other conditions are that the probabilities be nonnegative, that the sum of the probabilities over all alternatives equal one, that the probabilities depend only on differences in utilities, and that the cross derivatives of the probabilities with respect to the arguments be symmetric. See Börsch-Supan (1991), Daly and Zachary (1979), or McFadden (1981).

[3] Note that it is redundant to include $\theta_k \le U_{1k}(\bar{v})$ in the definition of f_2, since $U_{2k}(\bar{v}) \le U_{1k}(\bar{v})$. See Herriges and Kling.

[4] In the Bayesian framework, this can be interpreted as approximately a 13.5% probability that the first necessary condition is satisfied for all four modes.

A Reanalysis of the Empirical Evidence Regarding Willingness to Accept and Willingness to Pay and its Theoretical Consistency

Beatrice Lucero

Department of Agricultural and Natural Resource Econmics
Colorado State University and
Rocky Mountain Forest and Range Experiment Station
U.S.F.S., U.S.D.A.

Introduction

Willingness to accept (WTA) and willingness to pay (WTP) are theoretical constructs often used to measure, in monetary terms, welfare changes associated with changes in environmental quality through use of a survey technique called the contingent valuation method (CVM). When there are no quantity constraints, the difference between WTA and WTP is the income effect; WTP is constrained by budget constraints, but WTA is not. The theoretical appropriateness of either measure is determined by property rights. If society as a whole has property rights to environmental quality, then the theoretically appropriate money measure of welfare change associated with a reduction in environmental quality is WTA and for an increase, WTP.

Willig (1976) argued that under certain conditions, any difference between WTA and WTP is likely to be smaller than the probability of error associated with estimating the Marshallian demand curve, so using consumer surplus (CS) as a money measure of welfare change to approximate either WTA or WTP could be justified. Willig derived his results with respect to price changes. Randall and Stoll (1980) derived Willig's results with respect to quantity changes, such that the results are applicable to changes in environmental quality. The implications of the Willig and Randall and Stoll articles for natural resource valuation are that if there should not be a large difference between WTA and WTP, then it should not matter which measure is used. Pursuant to Willig and Randall and Stoll, researchers have sought to test the equality of WTP and WTA in the face of seemingly negligible or zero income effects. Empirical results (summarized in Table 1) have been reported as inconsistent with economic theory because WTA has consistently been significantly greater than WTP. (Fisher et al., 1988 and Cummings et al., 1986.)

The reported theoretical inconsistency along with empirical results of significant differences between hypothetical and actual WTP have contributed to skepticism over CVM's ability to evaluate welfare change. A panel appointed by the National Oceanic and Atmospheric Association (NOAA) to evaluate contingent valuation (Arrow et al., 1993) has recommended conducting CVM strictly in a WTP question format in spite of the theoretically appropriate measure of welfare change.

Table 1

Investigator	Good	WTA/WTP
Bishop & Heberlein (1985)	Goose hunting permit	$(4.8)_H$
Experiment 2	Deer hunting permit	$(19)_H$*; $(62)_R$**
Experiment 3	Deer hunting permit	$(16.5)_H$; $(6.88)_R$
Brookshire, et al. (1980)	Elk hunting permit:	
	.1, 1 elk encounter	$(1.57)_H$
	5 elk encounter	$(2.64)_H$
	10 elk encounter	$(6.47)_H$
Gordon & Knetsch (1979)	Fishing hole	$(\text{large})_H$
Gregory (1986)	Hand calculator	$(\text{significant})_R$
	Imported champagne	$(\text{significant})_R$
Hammack & Brown (1974)	Waterfowl hunting	$(4.23)_H$
Kahneman, et al. (1990)	Coffee mug	about $(2.33)_R$
	Ballpoint pens	about $(2.75)_R$
Knetsch & Sinden (1984)	lottery tickets:	
Experiment 1	\$70 in merchandise or \$50 in cash	$(\text{1/2 WTP group pd \$2; 76\% WTA group refused \$2})_R$
Experiment 2	\$15 in merchandise or \$10 in cash	$(\text{1/2 WTP group pd \$.50; 87\% WTA group refused \$.50})_R$
Experiment 3	WTP group given \$3; WTA group: \$70 in merchandise or \$50 in cash	$(\text{38\% of WTP group bought ticket; 82\% of WTA group refused to sell for \$3})_R$
Experiment 4	WTP group given \$1-\$4; WTA group: \$90 in merchandise or \$70 in cash	$(\text{20\% of WTP group bought ticket (75\% at lower price); 58\% of WTA group refused})_R$

* $(\text{WTA/WTP})_{\text{Hypothetical}}$, ** $(\text{WTA/WTP})_{\text{Real}}$

This paper presents a reanalysis of the empirical results summarized in Table 1; the results are determined to be theoretically consistent. The source of the income effects in the studies is identified as consumption at zero marginal cost (MC) (free experimental private goods or goods provided by the public sector at zero MC to individuals) or MC less than marginal utility (public goods provided at MC>0).

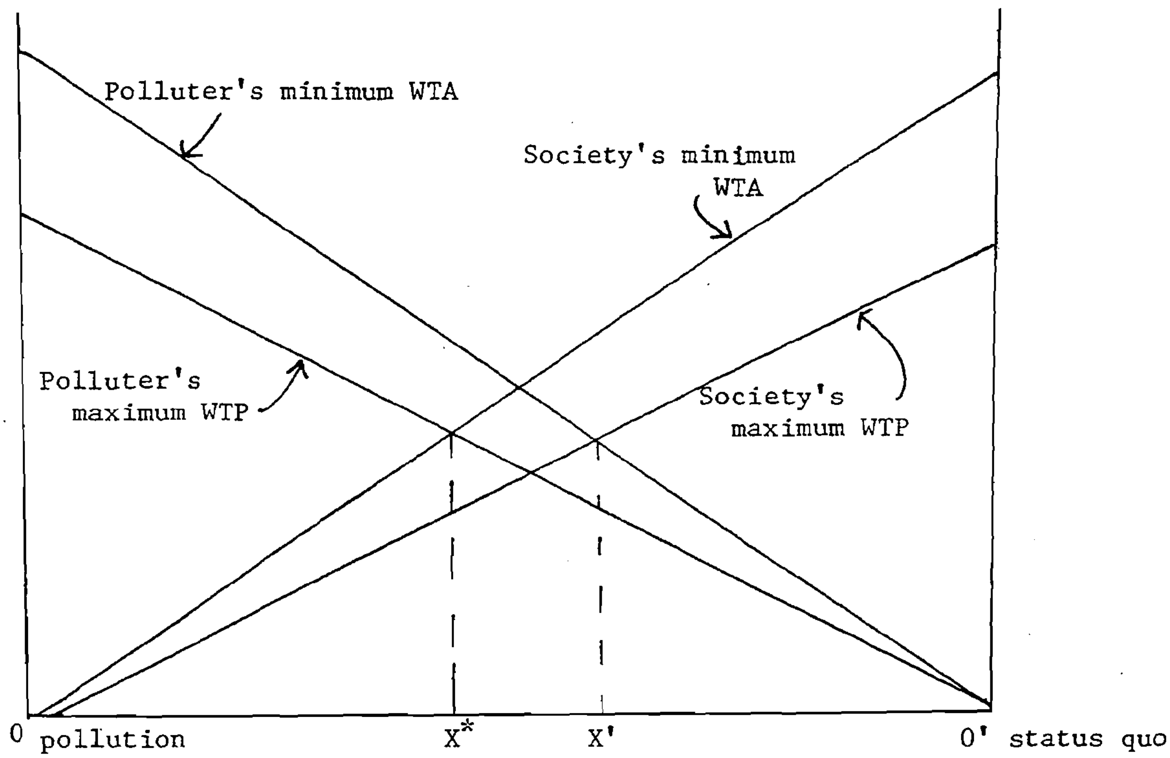


Figure 1.

The implications of the theoretical consistency established in this paper are that the property rights implied by a CVM question format will significantly affect the value placed on the subject public good. Suppose society has property rights to environmental quality. In Figure 1, negotiations would begin at 0 and proceed to the right. The intersection of society's WTA and the polluter's WTP would be relevant, so X^* would be the optimal level of pollution. On the other hand, if polluters have property rights to pollute, negotiations would begin at 0' and proceed to the left. The intersection of society's WTP and the polluter's WTA would be relevant, and X' would be the optimal level of pollution.

The significance of this paper is most relevant to cases where it is agreed that society has property rights to environmental quality and where the courts have established that property rights lie with society as a whole, in which case WTA would be the theoretically appropriate measure of welfare change from society's perspective. In these cases, if society is put in the position of not having property rights because of adherence to the NOAA panel recommendation, then in Figure 1, X' would be the resulting level of pollution, but X^* would be the socially optimal level of pollution. In order to frame CVM in a WTP format when WTA is theoretically appropriate, the survey must be designed such that respondents are asked what they are willing to pay to keep something that is already perceived as belonging to society. This is intended to serve as a proxy for the appropriate measure, willingness to accept compensation for a reduction in environmental quality. The question is much different from asking what they would be willing to accept as compensation for a loss in environmental quality and from asking what they would be willing to pay for an increment in environmental quality, because of the implied property rights. This appropriation of property rights from society is what is happening in natural resource valuation today; it is inefficient and unsustainable because environmental quality is being significantly under-valued, such that the values elicited by CVM do not reflect the relative scarcity of environmental quality.

In light of the empirical results, Knetsch (1990) has analyzed the implications of theoretically inappropriate use of WTP. He concludes that gains from trade may be overstated; final allocations will not be independent of property rights as purported by the Coase theorem, but rather can vary radically with property rights; loss assessments will be seriously understated, and decisions based on these assessments will be severely biased downward; too many environmentally destructive projects, and too little mitigation will be undertaken; welfare losses will not be adequately compensated; environmental control standards will be set too low; and benefit-cost analyses will be biased upward because costs will be underestimated.

In addition to the implications identified by Knetsch, lower income people will be under-represented in social welfare analyses because of the budget constraints imposed by a WTP framework, when WTA is the theoretically appropriate measure of welfare change. Further, if respondents perceive themselves as having property rights but are put in a position of not having property rights by a WTP question format, there is no apparent reason for believing the responses to be valid responses to the valuation questions asked. Strict adherence to the NOAA panel's recommendation will result in significant undervaluation of environmental quality and natural resources in cases in which society is deemed to have property rights such that WTA would be the theoretically appropriate measure. This would be the case for proposed reductions in environmental quality and for damage assessment cases. The courts have established that environmental property rights lie with society as a whole by holding polluters liable for damages; and there is no disagreement in the literature that WTA is the theoretically appropriate measure in environmental damage assessments. Yet WTP has become the prevalent measure of welfare change, as society continues to be faced with reductions in environmental quality.

The following section explains the difference between WTA and WTP in the private goods experiments within the scope of existing economic theory, demonstrating the theoretical consistency of the empirical findings. Income effects with respect to public goods are analyzed later in the paper, for the case of quantity changes.

Consistency with Microeconomic Theory - Private Goods Analysis.

The studies in Table 1 all involved goods that could be consumed at a MC lower than the individuals' marginal utility of the good. In the private goods experiments, the experimental good could be consumed at zero MC by the WTA groups. These experiments involved a WTA group and a WTP group. The WTA group was given a free good that was considered an insignificant contribution to consumer surplus, so it seemed reasonable to assume a negligible income effect. Therefore, according to Willig and Randall and Stoll, WTA and WTP would be close, or equal if there was zero income effect. The studies were consistent in finding WTA to be significantly greater than WTP, counter to Willig's and Randall and Stoll's predictions. Interestingly, the findings of WTA>WTP persist in real cash transactions.

Based upon the consumer's primary problem of maximizing utility, U(Q, X), subject to budget constraints, $M=P_QQ+P_XX$, (Q represents the private experimental good, X = all other goods, P_Q = the price of Q, P_X = the price of X, and M = money income), the Lagrange equation and first order conditions (FOC) for utility maximization are:

(1) $\mathcal{L} = U(Q, X) + \lambda(M-P_QQ-P_XX)$

(2) $\mathcal{L}_Q = U_Q - \lambda P_Q = 0$
(3) $\mathcal{L}_X = U_X - \lambda P_X = 0$
(4) $\mathcal{L}_\lambda = M - P_QQ - P_XX = 0$

The following relationship between the marginal utility of income (λ) and the income effect is derived from the second order conditions.

(5) $d\lambda/dP_Q + Q(d\lambda/dM) = -\lambda(dQ/dM)$

This equation can be found in Mishan (1981, pp. 523-524). It reveals that the income effect is weighted by λ. It is derived from a system of ordinal equations, so the findings are general.

The effect of zero MC on λ and rational consumer behavior.

Within the above system of equations, a definitive relationship between λ and P_Q cannot be determined; however, if cardinality is assumed, then one equation can be analyzed independently of the others, as in partial equilibrium analysis. Solving (2) for λ yields:

(6) $\lambda = U_Q/P_Q$

The marginal utility of income is defined as *marginal utility per dollar of cost*; it is inversely related with P_Q. By non-satiation U_Q cannot be zero. Thus by definition and from equation (6), λ is infinite when goods can be consumed at zero MC and positive marginal utility is derived. Since the income effect is weighted by λ, it is possible to have a huge income effect regardless of the size of dQ/dM, so long as dQ/dM $\neq$0. Therefore, the large income effects associated with consumption at zero MC in the private goods experiments are theoretically consistent.

Intuition behind respondents' behavior.

The effect of the receipt of a free good can be seen in Figure 2, which depicts the case where only one unit of Q can be consumed at zero MC. Assume that all of the individuals walking into the private goods experiments are at point A on U_0, where they all face market prices P_0. Receipt of a good at zero MC places the WTA individuals at point B on U_1, while the WTP individuals remain at point A. P_1 is a virtual price ratio. (Neary and Roberts, 1980.) The WTA group then faces the same opportunity costs as the WTP group, so the budget line for the WTA group will run parallel to the original budget line at market prices P_0 and through point D. However, the WTA individual's indifference curve is not tangent to the market price ratio at D. In fact, at D, P_0 is less than the marginal rate of substitution of Q for X ($MRS_{Q,X}$). If the WTA individuals were in the market for the good they would not sell at P_0. In some of the experiments (e.g. Kahneman) real mean WTA was less than the actual market price of the good, indicating that perhaps the WTA individuals were not in the market for the good. This is consistent with the assumption that they entered the experiment at point A, i.e. their demand for the good had already been satisfied. See Smith for a discussion on experimental results when respondents already own the good (1994, p. 141).

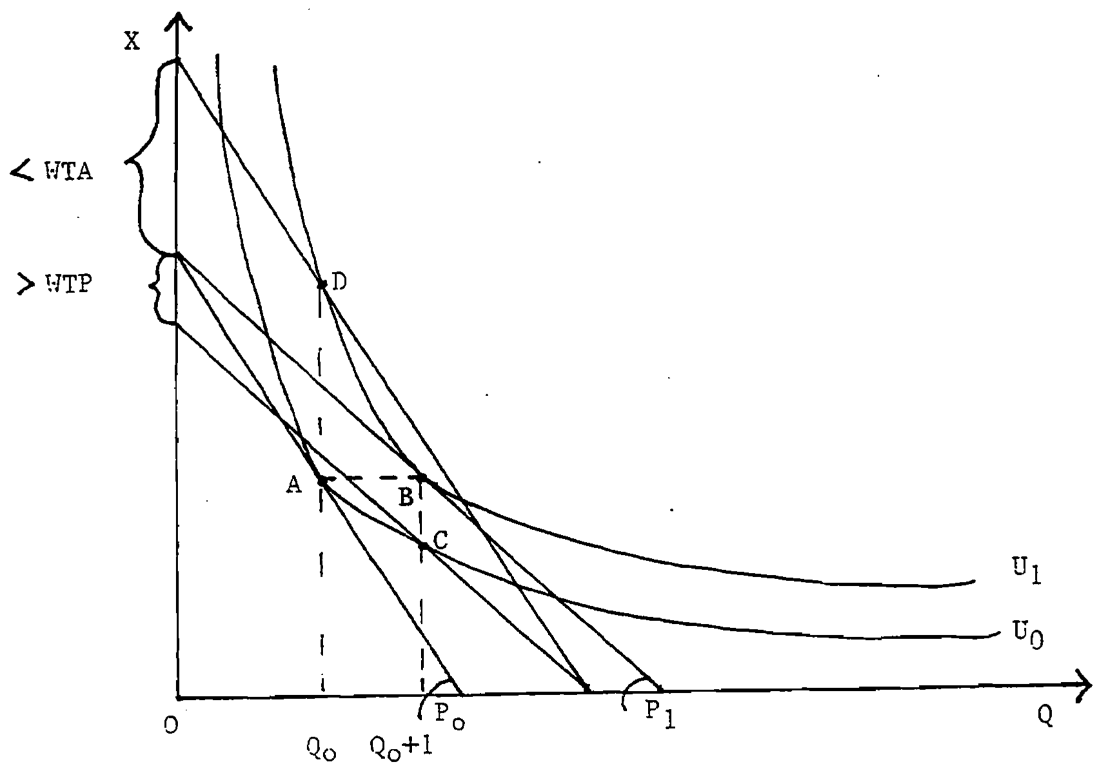


Figure 2.

At P_1, the WTP group's MRS_{QX} is less than P_1, so they would not be willing to purchase an additional unit of Q at P_1. They would be even less willing to pay P_0. A price ratio lower than P_0 and lower than P_1 would be necessary to induce the WTP group to purchase the experimental good; whereas a price ratio greater than P_0 and P_1 would be necessary to induce the WTA group to sell, if they were in the market for the good. That WTA>WTP when the WTA group receives goods at zero MC is consistent with rational utility maximizing behavior. The experiments have imposed the largest income effect possible indirectly by inducing an infinite λ and thereby a large income effect, by (5).

One comment on these results was that the initial MC_Q to the WTA group is not zero because they face the opportunity costs associated with retaining the good. These opportunity costs do not arise until after receipt of a free good and placement on a higher indifference curve. The fact that the WTA group then faces a positive opportunity cost is reflected in P_0.

Consistency with Microeconomic Theory - Public Goods Analysis

For the analysis of public goods, let Q = environmental quality. A distinguishing characteristic of public goods is that they can be consumed at either $MC<U_Q$ (scenario 1) or at MC=0 (scenario 2). This does not mean additional units of Q can be acquired at MC=0 or $U_Q>MC_Q$. It means that public goods **already in existence** can be consumed on a daily basis at MC=0 or $U_Q>MC_Q$, because of the non-rivalry and/or non-excludability characteristics of pure public goods. A non-rival good is one which does not diminish with consumption by any one individual, as long as demand for the good is below the level of congestion. The MC of providing a non-rival public good to another individual (once provided to any one individual) is zero. This is why it is inefficient to charge a positive price for non-rival goods, regardless of whether they are excludable. Non-excludability is the other defining characteristic of a pure public good. A good is non-excludable if, once it is provided, no one can be excluded from its consumption, such that it is not possible to enforce a positive price. A pure public good is both non-rival and non-excludable, but a public good can be characterized by any combination of these two characteristics. Since MC=0 or $U_Q>MC_Q$ for public goods already in existence, this analysis applies to CVM when WTA is the theoretically appropriate measure of welfare change. This is consistent with the private goods experiments where only the WTA groups enjoyed Q at MC=0.

Figure 3 demonstrates the way in which the optimal provision of public goods is theoretically determined. Individual demand curves are vertically summed, and the efficient quantity is Q^*. At Q^*, individual (1) would be willing to pay P_1, individual (2) would be willing to pay P_2, and so on. However, efficient pricing is not necessarily the government's objective, and perfect price discrimination is not technically feasible or desirable from an equity perspective. If provision of the public good is financed through taxes and price is set at P_Q, MC_Q would be less than the individual marginal valuations of Q (P_1, P_2, P_3, and P_4). If P_Q is the elk hunting license fee in the Brookshire et al. study, the marginal utility of hunting ($P_i=U_Q$) would be greater than P_Q. If $P_Q=0$ as in the Bishop and Heberlein geese and deer hunting studies, again $U_Q>P_Q$. If a less than socially optimal quantity such as Q' were provided at P_Q, the relationships would still hold and the differences between U_Q and P_Q

would be even greater. It is possible for U_Q to be less than P_Q, i.e. for individual 0 in Figure 3. The remainder of this analysis applies only to the case where $U_Q > P_Q$.

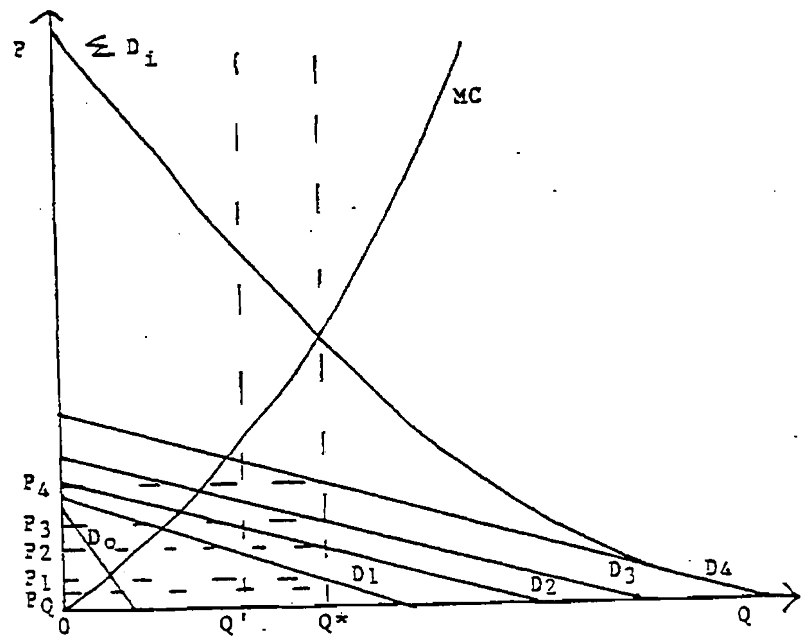


Figure 3.

The case where $MC_Q>0$.

Johansson (1987) has analyzed the consumer's utility maximization problem when Q is not a choice variable and the constraint on Q is binding under scenarios (1) and (2) as described above. He defines income in the budget constraint as income net of lump sum taxes used for the provision of public goods. For scenario (1) ($MC_Q>0$), Johansson derives the following set of equations (p. 58):

(7) $\partial V/\partial P_X = -\lambda X(P_X, M-P_Q, Q)$
(8) $\partial V/\partial M = \lambda(P_X, M-P_Q, Q)$
(9) $\partial V/\partial P_Q = -\lambda Q$
(10) $\partial V/\partial Q = \{\partial U[X(P_X, M-P_Q, Q), Q]\}/\partial Q - \lambda P_Q$

where V represents the indirect utility function, M is income net of lump sum taxes, and all other notation is the same as before. Equations (10) and (2) are similar: the FOC require that $U_Q=\lambda P_Q$. Assuming cardinality and solving for λ yields equation (6). The larger the difference between U_Q and P_Q, the larger will be λ and therefore the income effect. Individual (4) in Figure 3 would experience a larger income effect than the others.

According to Johansson, an increase in Q has both a substitution effect and an income effect. The income effect would cause the individual to reduce expenditures on X in order to consume more Q, since the constraint is binding. The substitution effect comes about because U_X is affected by changes in Q. An increase in P_Q has the income effect, $\partial V/\partial P_Q = -\lambda Q$, but no substitution effect, since an individual would be unwilling to reduce consumption of Q (binding constraint). (pp. 58, 59.) Again, λ is a significant determinant of the income effect.

The foregoing seems to imply that Q is a choice variable. Since $MC_Q>0$, these public goods are excludable, such that an individual need not consume Q. For Q to not be a choice variable means that individuals have no control over the level of Q accessible; but once Q is provided, they have a choice of how many times to consume Q, e.g. how often to visit the Grand Canyon.

The case where $MC_Q=0$.

With respect to scenario (2), Johansson states that there will be no income effect associated with a change in Q since $P_Q=0$. The income effect is that part of a price change that forces an individual on to a different indifference curve because it necessarily changes real income. This can be seen in Figure 4 with respect to quantity changes.

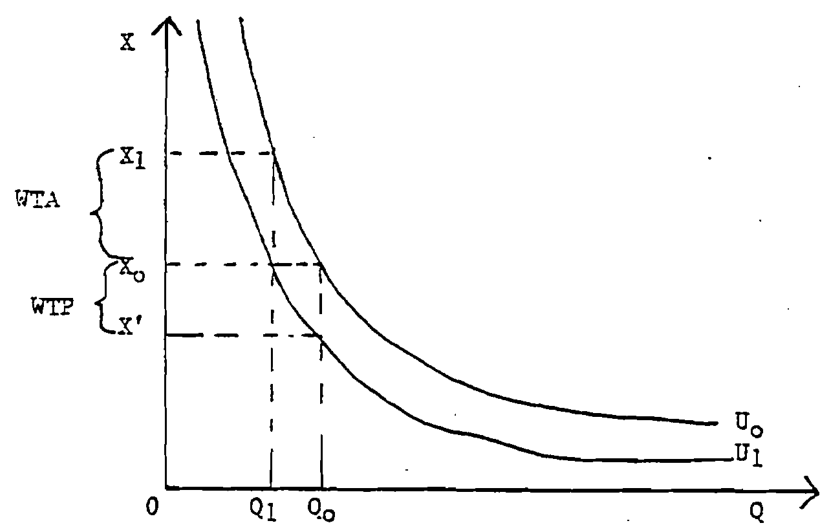


Figure 4.

A reduction in Q from Q_0 to Q_1 will force the individual on to a lower indifference curve, from U_0 to U_1. In order for the individual to remain on U_0 he or she must be compensated for the income effect by the amount (X_1-X_0). If $P_Q=0$, but $P_X\neq0$, there can be no substitution effect because X cannot be consumed at zero MC. A reduction in the quantity of Q is equivalent to an increase in the virtual price of Q. In Figure 2, it would be represented by a move from point B to point A. Recall from Johansson's analysis, that a change in P_Q has an income effect but no substitution effect, because the constraint on Q is binding.

Conditions under which income effects associated with a reduction in Q can be zero.

(1) The case where U_Q=0.

In order for there to be no income effect associated with a reduction in Q, it must be the case that the individual can remain on the same indifference curve. One way for this to be the case is for U_Q to be zero. This case is depicted in Figure 5:

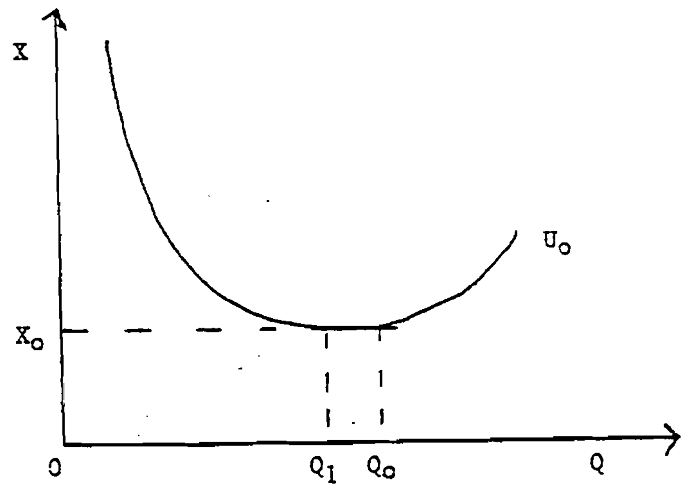


Figure 5.

Johansson's conclusion of zero income effect relies upon U_Q=0 by equation (10). However, P_Q has been set below U_Q; and by non-satiation U_Q cannot be zero, such that equality in (10) cannot hold. It **cannot** be assumed that U_Q=P_Q as in private goods analyses.

The expenditure function is derived from V and assumes that (10) has been met with equality. Findings of zero income effect via expenditure function analyses also rely on U_Q=0. Expenditure functions are based upon money income (when real income is relevant) and cannot incorporate the fact that for public goods U_Q>P_Q. Freeman's conclusion (1993, p. 74) that income effects are necessarily zero when P_Q=0 is based on expenditure functions. He states that "If w_q is the marginal value of a change in q [environmental quality], it is given by the derivative of the restricted expenditure function with respect to:

$$w_q = -\partial e/\partial q$$

The right-hand side of this expression is also equal (in absolute value) to the slope of the indifference curve through the point at which welfare change is being evaluated." This is the case depicted in Figure 5. The implication is that since environmental quality is unpriced ($-\partial e/\partial q=0$), its marginal value is zero.

(2) The case where λ is held constant with respect to all prices.

Another way for it to be the case that there is no income effect with respect to quantity changes when $P_Q=0$ is to hold λ constant with respect to all prices (including virtual prices), which is what Willig does. By focusing on the income elasticity of demand to measure income effects, and then by holding that constant at 1; or such that the smallest and largest values of the income elasticity of demand (η^o and η') are "sufficiently close in value" (Willig, p. 594.), it is clear that Willig's findings are based upon Samuelson's first interpretation of λ being constant. Prices are allowed to vary, but they are assumed to have no effect on λ. "Thus, if $\lambda = \lambda[(M)]$, the money measure (consumer surplus) gives an exact or at least proportional measure of utility change, i.e., $S=\Delta U/\lambda$, when prices vary with income fixed," where S = consumer surplus (Johansson, p. 28.)

Willig states that his analysis does not depend on a constant λ. This appears to be based upon his assertion that his findings hold even if all income elasticities of demand are not assumed to be 1, as implied by homothetic utility functions and λ's homogeneity of degree minus one. Willig (p. 592) defines constant income elasticity of demand:

$$[\partial X_1(p,M)/\partial M][M/X_1(p,M)] \equiv \eta$$

Willig assumes that there is no difference between η^o and η'. When he relaxes this assumption, he still requires that η^o and η' be "sufficiently close in value." (p. 594.) He asserts that "(M)easured income elasticities of demand tend to cluster closely about 1.0, with only rare outliers." (p. 590.) Hanemann (1994) reports income elasticities of demand for state and local government services as ranging from .3 to .6, for charitable giving ranging from .4 to .8, and in CVM falling within a similar range. McKenzie (1983) also claims that there is no empirical justification for asserting that income elasticities of demand cluster about 1.0 and further states that:

> if (η^o and η') are assumed to be close in value, it must also be the case that the range of price variation under consideration is not very great. Consequently, it will be true that the magnitude of the consumer surplus integral will be small. (p. 118.)

McKenzie uses an example based upon the Klein-Rubin linear expenditure system and a difference between η^o and η' of .20 and determines that in order to generate Willig's results "the maximum allowable range of price variation is -.002" (p. 118).

From the derivation of Willig's results presented in Just et al. (pp. 97-103), it is clear that without assuming a small $|\Delta P|$, then it cannot be assumed that $|\Delta P|X$ = consumer surplus, or that a change in consumer surplus is equal to the real income change caused by ΔP. Further, if the demand curves are not linear, then the areas between WTA and CS and between CS and WTP will not be triangles. Johansson reports that Willig's results were derived for a single consumer, and that if "there are large variations in income and/or income elasticity of demand between consumers, the aggregate error may become quite large" (p. 53).

Willig may or may not be holding $\partial X/\partial M$ constant, but he is holding $\partial \lambda/\partial P_i$ constant. Since the objective of CVM is to sum individual measures of welfare change, we know that homothetic utility functions are in fact assumed. Willig acknowledges that his results apply in the case of homothetic utility functions (McKenzie). Given Samuelson's first interpretation of the constancy of λ, we know that homothetic utility functions necessarily imply that λ is held constant with respect to price changes, so there is no question that Willig is holding λ constant with respect to prices. Thus Willig's findings rely on a constant λ with respect to all prices by assumption of marginal price changes. His findings apply only when income effects are zero or negligible - since in his analysis income effects are negligible or zero by design.

(3) The case where λ is held constant with respect to (n-1) prices and income.

Another way for it to be the case that there is no income effect associated with quantity changes when $P_Q=0$ is to hold λ constant with respect to (n-1) prices and income. When Freeman concludes (in a separate analysis from that described above) that there is necessarily no income effect when $P_Q=0$, he assumes quasi-linear utility functions. This can be seen in his Figure 3.10 (p. 78). Again, income effects are zero by design.

Conclusion

The empirical findings of WTA>WTP with respect to the private experimental goods are theoretically consistent because when goods can be consumed at zero MC, λ is infinite by definition, and the income effect is weighted by λ. If researchers really want to test for the equality of WTA and WTP with respect to private goods, the experiments need to be designed such that participants are not given free goods and are known for certain to be in the market for the experimental good. If WTA>WTP persists under these conditions, then theoretical consistency can be questioned, assuming an otherwise appropriate experimental design.

The same conclusion holds with respect to public goods that are provided at zero MC to individuals, so long as people disassociate their taxes with the provision and consumption of environmental quality, as is explicitly assumed by Johansson and implicitly by Freeman. With respect to public goods that are provided at some positive MC to consumers, the larger the difference between MC_Q and U_Q, the larger will be λ and therefore the income effect. Economists have relied upon income elasticities of demand in determining income effects, thereby attributing the entire income effect to dQ/dM, even though the income effect is weighted by λ.

Economic analysis is incomplete with respect to income effects associated with consumption at zero MC and $MC_Q<U_Q$. For there to be no income effect associated with changes in Q simply because $MC_Q=0$ requires either assuming $U_Q=0$, holding λ constant with respect to all prices, or holding λ constant with respect to (n-1) prices and income, neither of which is deemed justified by a review of the literature. To argue for equality of WTA and

WTP for public goods fails to recognize the relationships between binding constraints on Q, consumption at $MC_Q<U_Q$, $\lambda=U_Q/MC_Q$, and the weighting of the income effect by λ.

Willig's analysis is based upon income elasticities of demand. Brookshire et al.'s and Randall and Stoll's is based upon the price flexibility of income which is a function of the income elasticity of demand and the elasticity of substitution. Freeman's analysis is based upon the expenditure function. Whereas all of these measures are based upon money income, it is real income that is relevant. If $P_Q=0$ and $U_Q>0$, a reduction in Q represents a fall in real income because the initial level of money income now acquires a bundle with less Q in it; and an increase in Q represents an increase in real income because the initial level of money income now acquires a bundle with more Q in it. These are income effects by definition, since real income changes, and individuals are forced on to different indifference curves.

Equation (11) reflects the indirect relationship between money and utility and makes it clear that it is the consumption of goods that yields utility and not money itself.

(11) $\lambda = (\partial U/\partial Y)(\partial Y/\partial M)$

where Y represents a vector of all goods. Money yields utility only to the extent that it can be used to acquire goods for which $U_Y>0$. If utility can be derived at MC=0, money is not a necessary condition for utility derivation. From the consumer's perspective, there is no better way to maximize utility while minimizing costs than to consume goods at zero MC.

Marginal analysis is appropriate for components of environmental quality that are continuous e.g. air quality and water quality, but with respect to discrete quantity changes, it is undefined. The above analysis nevertheless demonstrates the source of and size of income effects associated with consumption of public goods.

Hanemann (1991) found that large differences between WTA and WTP are theoretically consistent for a good that has few or no substitutes. For quantity constraints, differences between WTA and WTP also include a substitution effect. A graphical analysis can be found in Just et al. (pp. 136-142). The analysis in this paper implies that large differences between WTA and WTP for public goods are theoretically consistent even if there are substitutes because public goods can be consumed at $MC_Q<U_Q$ and sometimes at MC=0.

Bibliography

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. "Report of the NOAA Panel on CV(M)." (1993, January 11.) Resources for the Future.

Bishop, Richard C. and Thomas A. Heberlein. (1985, Oct 21-16.) "Assessing the Validity of Contingent Valuation: Three Field Experiments." Paper presented at the International Conference, Man's Role in Changing the Global Environment. Italy.

Bishop, Richard C. and Thomas A. Heberlein. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics. (1979, Dec.) 61(5):926-930.

Brookshire, David S., Alan Randall, and John R. Stoll. "Valuing Increments and Decrements in Natural Resource Service Flows." American Journal of Agricultural Economics. (1980, August.)
62(3):478-488.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze, Eds. (1986.) Valuing Environmental Goods: *an Assessment of the Contingent Valuation Method*. Rowman & Allanheld, New Jersey.

Fisher Ann, Gary H. McClelland, and William D. Schulze. "Measures of Willingness to Pay Versus Willingness to Accept: Evidence, Explanations, and Potential Reconciliation" in Amenity Resource Valuation, Integrating Economics with Other Disciplines. (Peterson, George L., B. L. Driver, and Robin Gregory, Eds.) (1988.):127-134.

Freeman, A. Myrick, III. (1993.) The Measurement of Environmental and Resource Values, Theory and Methods. Resources for the Future.
Gordon, Irene M. and Jack L. Knetsch. "Consumer's Surplus Measures and he Evaluation of Resources." Land Economics. (1979, Feb.) 55(1):1-10.

Gregory, Robin. "Interpreting Measures of Economic Loss: Evidence from Contingent Valuation and Experimental Studies." Journal of Environmental Economics and Management. (1986.) 13:325-337.

Hammack, Judd and Gardner Mallard Brown, Jr. (1974.) Waterfowl and Wetlands: Toward Bioeconomic Analysis. Resources for the Futures, Inc., Washington, D.C.

Hanemann, W. Michael. "Valuing the Environment Through Contingent Valuation." Journal of Economic Perspectives. (1994, Fall.) 8(4):19-43.

Hanemann, W. Michael. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" The American Economic Review. (1991, June.) 81(3): 635-647.

Hicks, J. R. (1981.) Wealth and Welfare. Camelot Press, Ltd., Southampton, Great Britain.

Hicks, J. R. (1957.) Value and Capital. Oxford University Press. 2d. Ed.

Hicks, J. R. (1959.) A Revision of Demand Theory. Oxford University Press. 2d. Ed.

Johansson, Per-Olov. (1987.) The Economic Theory and Measurement of Environmental Benefits. Cambridge University Press.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. (1982.) Applied Welfare Economics and Public Policy. Prentice-Hall, Inc.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Experimental Tests of the Endowment Effect and the Coase Theorem." Journal of Political Economy. (1990, December.) 98(6):1325-1348.

Knetsch, Jack L. and J. A. Sinden. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." Quarterly Journal of Economics. (1984, August.) 99(3):507-521.

McKenzie, George W. (1983.) Measuring Economic Welfare: New Methods. Cambridge University Press.

Mishan, E. J. (1981.) Introduction to Normative Economics. Oxford University Press.

Mitchell, Robert Cameron and Richard T. Carson. (1989.) Using Surveys to Value Public Goods, The Contingent Valuation Method. Resources for the Future.

Neary, J. P. and K. W. S. Roberts. "The Theory of Household Behavior Under Rationing." European Economic Review. (1980) 13:25-42.

Neill, Helen R., Ronald G. Cummings, Philip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin. "Hypothetical Surveys and Economic Commitments." Land Economics. (1994, May.) 70(2):145-154.

Nicholson, Walter. (1992.) Microeconomic Theory, Basic Principles and Extensions. The Dryden Press. 5th Ed. p. 135.

Randall, Alan and John R. Stoll. "Consumer's Surplus in Commodity Space." The American Economic Review. (1980, June.) 70(3): 449-455.

Shogren, Jason F., Seung Y. Shin, Dermot J. Hayes, and James B. Kliebenstein. "Resolving Differences in Willingness to Pay and Willingness to Accept." The American Economic Review. (1994, March.) 84(1):255-270.

Smith, V. Kerry. "Lightning Rods, Dart Boards, and Contingent Valuation." Natural Resources Journal. (1994, Winter.) 34(1): 121-152.

Tresch, Richard W. (1981.) "Consumption Externalities." In Public Finance *a Normative Theory* (pp. 93-139). Business Publications, Inc. Plano, TX.

Willig, Robert D. "Consumer's Surplus Without Apology." The American Economic Review. (1976, September.) 66(4):589-597.

Recovering Weakly Neutral Preferences

Douglas M. Larson*

Department of Agricultural Economics
University of California
Davis, CA 95616

Abstract

This paper demonstrates a method for recovering weakly neutral expenditure and indirect utility functions while integrating back from Marshallian demand specifications. A comparison with quasi-expenditures derived under weak complementarity for the same demand specifications shows that weak neutrality and weak complementarity are not equivalent assumptions. Weak complementarity can be seen to be nested within weakly neutral models, and can be tested for in appropriately formulated models.

Recovering Weakly Neutral Preferences

I. Introduction

The valuation of environmental amenities has grown in importance to policymakers (and interest to the economics profession) over the last decade, aided in large part by such high profile natural resource damage cases as the *Exxon Valdez* and *Montrose* cases. Among the developments which followed the *Exxon Valdez* case was the formation, by the National Oceanic and Atmospheric Administration, of a blue-ribbon panel of experts in survey research and economics to evaluate whether the contingent valuation method could be used in assessing lost passive use values from such cases. The panel, whose origins, purposes, and importance are well described by Portney, issued a qualified endorsement of the use of contingent valuation as a starting point to determine natural resource damages, including lost passive use value (Arrow *et al.* [1]). It also encouraged the development of alternatives to contingent valuation where practicable.

The scope of work aimed at valuing environmental amenities or natural resource damages by methods other than contingent valuation has been extremely limited. A primary reason is that in nationally-prominent damage or valuation cases, the passive use component of benefits can be large relative to the use values at issue. In contrast, the primary area where environmental economists have used methods other than contingent valuation is in measuring use-related benefits of amenity changes, most often in outdoor recreation (e.g., Bockstael and McConnell [2]) but not exclusively so (e.g., Dickie and Gerking [3]). It is widely recognized that when valuing changes in amenities,[1] some unobservable restriction on preference is required in order to use the individual's behavior (e.g., demand functions for related private goods) to infer valuations.[2] The most familiar restriction is weak complementarity between a set of private goods and an amenity of interest (Mäler [10]), which asserts that when a set of private market goods related to the amenity are not consumed, there is no change in the individual's utility when the amenity changes. Because invoking this preference restriction is tantamount to asserting there is no passive use value, economists have used weak complementarity in the course of assessing changes in welfare associated with amenity changes, with the consequence that all of the value so measured is, definitionally, use value. While some economists would argue that weak complementarity describes a public-private good link for which one can form some intuition ([2]), this view is not universally shared (LaFrance [6]).

As a consequence, then, of the fact that high profile damage cases are thought, potentially, to involve passive use values, and the fact that the only method used to date for analyzing the value of amenity changes through consumption of related private goods demands rules out passive use values, it is not surprising that there has been no empirical measurement by methods other than contingent valuation.

Recently, some attention has been given to using demand systems for amenity valuation with preference restrictions other than weak complementarity. Neill [13] showed that if sets of goods can be classified as Hicksian complements or substitutes for a non-market amenity, bounds on the marginal willingness to pay for the amenity could be derived. Larson [9] expanded on this theme, showing that with such information exact expressions, rather than bounds, could be derived.

Larson showed that a weaker form of this restriction, holding only at a given price vector rather than everywhere, was also sufficient to identify marginal amenity values. He argued that with this price vector chosen to reflect choke prices for goods that generate use value, "weak neutrality" could serve as an alternative preference restriction to weak complementarity in amenity valuations since it preserves Mäler's [10] basic insight about public and private good linkages but does not impose the assumption of zero passive use value *a priori*. This has not yet been incorporated into empirical work because neither the implications of weakly neutral (as opposed to weakly complementary) quasi-preferences derived from empirical demand equations, nor the methods for exact welfare measurement, are well understood.

A primary purpose of this paper is to help fill this gap by demonstrating how one can recover quasi-preferences corresponding to empirical demand functions under weak neutrality. The approach is similar to that taken in [8], which demonstated the process by which one could recover the quasi-expenditure function corresponding to weak complementarity. Given the ability to recover quasi-expenditure functions under weak neutrality, it is possible in principle to test whether weak complementarity holds as a special case within the weak neutrality framework. For such tests to be meaningful, the empirical demand specification must be sufficiently flexible to accomodate separate parameters representing use and potential passive use value. These issues are discussed in the context of a linear empirical demand specification.

A subsidiary purpose of the paper is to clear up some confusion that has arisen in a paper by Flores [4] on the subject of weak neutrality. That paper corrects a notational error in [9], and discusses how one can numerically approximate welfare effects of amenity changes under Hicks neutrality, developing expressions for the bias one may incur from assuming weak neutrality when it does not in fact hold. Flores also asserts incorrectly that weak neutrality is equivalent to weak complementarity. Given the quasi-preferences recovered from empirical demand functions under each of the alternative hypotheses, a comparison readily shows that the assertion is false.

I. Background

The same basic setup as in [8] is used. The primal consumer problem is the maximization of the utility function $u(\mathbf{x},z,q)$, where $\mathbf{x}$ is a vector of consumption goods with corresponding price vector $\mathbf{p}$, q is a scalar quality variable, and z is a scalar composite commodity such that $z=m-\mathbf{px}$, where m is income. The solution to this problem is the set of Marshallian demands $\mathbf{x}=\mathbf{x}(\mathbf{p},q,m)$ and $z=z(\mathbf{p},q,m)=m-\mathbf{px}$. Substituting these demands into the utility function yields the indirect utility function $v(\mathbf{p},q,m) \equiv u(\mathbf{x}(\mathbf{p},q,m),z(\mathbf{p},q,m),q)+\lambda[m-\mathbf{px}-z]$, where λ is a Lagrange multiplier. The inverse of indirect utility with respect to the income argument is the minimum expenditure function $e(\mathbf{p},q,u) \equiv \min_{\mathbf{x},z}\{\mathbf{px}+z \mid u(\mathbf{x},z,q)=u\}$. As expenditure varies to hold utility constant and compensate for changes in any price p_i, one can write

$$dv = v_i dp_i + v_m de(\mathbf{p},q,u) \equiv 0$$

which can be rewritten as the ordinary differential equation

$$\frac{de(\mathbf{p},q,u)}{dp_i} = -\frac{v_i}{v_m} = x_i(\mathbf{p},q,e(\mathbf{p},q,u)) \tag{1}$$

with the latter equality resulting from Roy's Identity. Integrating (1) sequentially for all p_i obtains the quasi-expenditure function $\tilde{e}(\mathbf{p},q,\theta(q,u))$ (see, e.g., Hausman [5]; LaFrance and Hanemann [7]), which is related to the true expenditure function by

$$e(\mathbf{p},q,u) \equiv \tilde{e}(\mathbf{p},q,\theta(q,u)) \tag{2}$$

where $\tilde{e}(\cdot)$ is a known function that represents the part of the expenditure function which is identified parametrically from (1), and $\theta(\cdot)$ is the unknown constant of integration. The quasi-expenditure function $\tilde{e}(p,q,\theta(q,u))$ contains all the information necessary for measuring compensating variations of changes in $\mathbf{p}$, but not for q without additional structure to identify $\theta(q,u)$.

Larson [8] showed how requiring preferences to exhibit weakly complementary provides sufficient additional structure for the constant of integration. Weak neutrality also provides a sufficient, albeit different, structure for identifying $\theta(q,u)$. The weak neutrality assumption follows from the assertion that the set of goods generating use value can be identified.[3] Under weak neutrality, $\mathbf{x}$ represents the good(s) which generate use value, and the Hicksian composite good $z(\mathbf{p},q,u)$ is composed of goods which do not generate use value. By definition of use value, then, at the choke vector $\hat{\mathbf{p}}$ for which $\mathbf{x}(\hat{\mathbf{p}},q,u) \equiv \mathbf{0}$, $\partial z(\hat{\mathbf{p}},q,u)/\partial q \equiv 0$ ([9]). From the Slutsky-Hicks equation governing the change in $z(\cdot)$ with q, with the budget constraint substituted in,

$$\partial z(\hat{\mathbf{p}},q,u)/\partial q = [-\sum_i \hat{p}_i \cdot \partial x_i(\hat{\mathbf{p}},q,e(\hat{\mathbf{p}},q,u))/\partial q]$$

$$+ [1 - \sum_i \hat{p}_i \cdot \partial x_i(\hat{\mathbf{p}},q,e(\hat{\mathbf{p}},q,u))/\partial m](de(\hat{\mathbf{p}},q,u)/dq) \equiv 0$$

so the operational expression for the marginal value of the amenity changes is

$$de(\hat{\mathbf{p}},q,u)/dq = [\sum_i \hat{p}_i \cdot \partial x_i(\hat{\mathbf{p}},q,e(\hat{\mathbf{p}},q,u))/\partial q]/[1 - \sum_i \hat{p}_i \cdot \partial x_i(\hat{\mathbf{p}},q,e(\hat{\mathbf{p}},q,u))/\partial m]. \qquad (3)$$

The expression comparable to (3) under weak complementarity is $de(\hat{\mathbf{p}},q,u)/dq = 0$ ([8], equation (5)). In order to derive expressions for the quasi-expenditure function under weak neutrality that are comparable with those derived previously under weak complementarity, $\mathbf{x}$ is taken to be a scalar representing consumption of a single good weakly neutral to the public good q.[4] In this case, p can be used unambiguously to represent price of the market good x weakly neutral to q. After substituting in the quasi-expenditure function from (2), (3) simplifies to

$$\frac{d\tilde{m}}{dq} = \frac{\partial \tilde{m}}{\partial q} + \frac{\partial \tilde{m}}{\partial \theta}\,\frac{\partial \theta(q,u)}{\partial q}$$

$$= [\hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial q]/[1-\hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial m]. \quad (4)$$

where $\tilde{m} \equiv \tilde{e}(\hat{p},q,\theta(q,u))$ is the quasi-expenditure function evaluated at the choke price $\hat{p}$. Equation (4) is the expression for marginal passive use value, because it describes the change in the expenditure function given that the related private good x is not being consumed.

As noted above, the structure of $\tilde{m}$ is known from integrating back over price, and $x(\cdot)$ is the Marshallian demand function which is also observable, so (4) is an ordinary differential equation in $\theta(q,u)$ and q which, if solvable, recovers the constant of integration, and hence the quasi-expenditure function, in terms of observables. Equation (4) is the weak neutrality condition for a single-equation demand system, which differs from the weakly complementary condition of [8] in that marginal passive use value, the expression on the right side of (4), is not set to zero by assumption.

II. Weakly Neutral Quasi-Expenditure Function for Linear Demand

If the Marshallian demand function for a good x of interest is of the form

$$x = \alpha+\beta p+\gamma q+\delta m \quad (5)$$

the corresponding quasi-expenditure function is

$$\tilde{e}(p,q,\theta(q,u)) = \theta(q,u)e^{\delta p}-(1/\delta)[\alpha+\beta p+\gamma q+\beta/\delta] \quad (6)$$

where $\theta(q,u) < 0$ is the constant of integration that in general depends on q.[5] The Hicksian choke price

$$\hat{p}(q,u) = (1/\delta)\ln\left\{\frac{\beta}{\delta^2\theta(q,u)}\right\}. \quad (7)$$

is the same as in the weakly complementary case. Using (7) in (6) gives the quasi-expenditure function evaluated at $\hat{p}$, which is

$$\tilde{m} = \Big\{\beta/\delta^2 - (1/\delta)[\alpha + \beta/\delta \ln\{\beta/\delta^2\theta(q,u)\} + \gamma q + \beta/\delta]\Big\}$$

$$= \Big\{-(1/\delta)[\alpha + \beta/\delta \ln\{\beta/\delta^2\theta(q,u)\} + \gamma q]\Big\}$$

from which the expression for the left side of (4) is

$$\frac{d\tilde{m}}{dq} = -\gamma/\delta + (\beta/\delta^2)\frac{d\theta/dq}{\theta}, \tag{8}$$

since $\partial\theta/\partial q = d\theta/dq$. The expression for the right side of (4) is obtained by substituting both (6) and (7) in (5), yielding $x(\hat{p},q,\tilde{m})$, the ordinary demand evaluated at the Hicksian choke price with income allowed to vary according to the quasi-expenditure function. The weak neutrality condition given on the right side of (4) is then

$$[\hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial q]/[1 - \hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial m] = \frac{\gamma\cdot(1/\delta)\ln(\beta/\delta^2\theta)}{1 - \delta\cdot(1/\delta)\ln(\beta/\delta^2\theta)}$$

$$= \frac{(\gamma/\delta)\ln(\beta/\delta^2\theta)}{1 - \ln(\beta/\delta^2\theta)}. \tag{9}$$

Equating (8) and (9) gives the weak neutrality condition for the linear demand model, which is an ordinary differential equation in $\theta(q)$ and q; gathering terms and simplifying, it can be written

$$-\gamma/\delta + (\beta/\delta^2)[1 - \ln(\beta/\delta^2\theta)]\frac{d\theta/dq}{\theta} = 0. \tag{10}$$

This differential equation is exact, and integrates back to a function $F(\theta(q),q)=K$, with K some constant that can be scaled as the utility index. Taking $\partial F/\partial q = -\gamma/\delta$ and $\partial F/\partial\theta = (\beta/\delta^2)[1 - \ln(\beta/\delta^2\theta)]/\theta$ and integrating, one obtains

$$F(\theta,q) = \ln(-\theta) + \tfrac{1}{2}[\ln(\beta/\delta^2\theta)]^2 - (\delta\gamma/\beta)q = \ln(-\phi) \tag{11}$$

after scaling by the constant δ^2/β.[6] This implicit function representation of the constant of integration is useful for comparison with the constant of integration derived under weak complementarity, which is $\theta = \phi\cdot\exp[(\gamma\delta/\beta)q]$ ([8], p. 102), or (recalling that θ and ϕ are both negative)

$$\ln(-\theta) - (\delta\gamma/\beta)q = \ln(-\phi) \tag{12}$$

Two points are important about the comparison of equations (11) and (12). First, it indicates quite clearly that the constant of integration, and therefore the quasi-expenditure function, is different under weak neutrality than under weak complementarity, disproving the claim in [4]. Second, and more importantly, it shows the possibility of testing for weak complementarity within the weak neutrality framework, since (11) and (12) are equivalent if and only if $\beta/\delta^2\theta = 1$. This point is taken up further in Section III.

Using the quadratic formula to solve (11) explicitly for $\theta(q,u)$, and substituting into (6), one obtains the quasi-expenditure function under weak neutrality consistent with the linear demand in (5). This is

$$\tilde{e}(p,q,u) = (\beta/\delta^2)e^{-1+\sqrt{1+2[(\delta\gamma/\beta)q-u]}+\delta p} - (1/\delta)[\alpha+\beta p+\gamma q+\beta/\delta] \tag{13}$$

with the positive root chosen so that $\partial\tilde{e}/\partial u > 0$, consistent with the requirements of theory. By contrast, the weakly complementary quasi-expenditure function consistent with equation (5) ([8], p. 102) can be written as

$$\tilde{e}(p,q,u) = (\beta/\delta^2)e^{(\delta\gamma/\beta)q - u + \delta p} - (1/\delta)[\alpha+\beta p+\gamma q+\beta/\delta]. \tag{14}$$

III. Testing for Weak Complementarity in the Weak Neutrality Framework

Section II established that the quasi-expenditure functions obtained under weak neutrality and weak complementarity are in general different. Given that each is parameterized in terms of observable parameters of Marshallian demand functions, one can in principle test whether weak complementarity holds as a special case of weak neutrality. Since weak complementarity implies that passive use value is identically zero (while weak neutrality does not), this is equivalent to a test whether the quasi-expenditure function recovered from a given demand specification reflects only use value or a combination of use and passive use value.

Such tests could be performed on the weakly neutral quasi-expenditure function recovered in Sections II. However, the empirical demand specification which gave rise to that quasi-expenditure function is too simple and inflexible to yield meaningful tests, because the amenity values (whether use or passive use) are tied to a single demand parameter. A more appropriate specification would allow the empirical specification to accomodate either use or passive use value, or both. When these amenity values are reflected in more than one parameter estimate, more meaningful tests may emerge.

These points can be seen in the linear demand specification in Section II. It was previously noted following equation (12) that if $\beta/\delta^2\theta = 1$, the ordinary differential equations for θ and q are equivalent under both weak complementarity and weak neutrality.[7] Since β and δ are demand parameters whose estimates have a variance-covariance matrix in empirical applications, one might consider a test of the form H_0: $\theta = \beta/\delta^2$ versus the alternative H_1: $\theta \neq \beta/\delta^2$.[8] However, the implications using this functional form are extreme: from (7), if $\theta = \beta/\delta^2$, the choke price $\hat{p} = 0$. Furthermore, from (11) and (12), the only way that $\theta=\beta/\delta^2$ can hold for general amenity level q and ϕ a constant is for $\gamma = 0$, meaning the level of the amenity has no effect on Marshallian (and Hicksian) demand.

This specification is uninteresting for the purposes of testing for weak complementarity as a special case of weak neutrality it implies that both use and nonuse value are zero. This results because the specification is too parsimonious with respect to how the amenity level affects demand. To allow for use and passive use to be reflected separately in the quasi-expenditure function, the demand specification must have at least two parameters reflecting the effect of the amenity on behavior. A more promising specification might be

$$x = \alpha+\beta p+\gamma_1 q+\gamma_2(\hat{p}-p)q+\delta m \tag{15}$$

which allows for different effects of the amenity on use and passive use value. If the related private good is being consumed ($p < \hat{p}$), the marginal effect of the amenity on Marshallian demand is reflected through two parameters, $\gamma_1+\gamma_2(\hat{p}-p)$; if, on the other hand, the private good is not being consumed ($p \equiv \hat{p}$), which defines the conditions under which change in the (quasi-) expenditure function is passive use value, the term involving γ_2 drops out and the marginal effect of changes in the amenity are the same as in the model of Section II, resulting from the γ_1 parameter. One would expect, then, that a test of whether passive use value is a significant part of the amenity value can be formulated in terms of γ_1, while a test of the significance of use value would involve both γ_1 and γ_2.

To see how these tests can be developed from the structure of (15), integrating back over price in the same manner as in Section II (following [8] and the references cited therein) yields the quasi-expenditure function

$$\tilde{e}(p,q,\theta(q,u)) = \theta(q,u)e^{\delta p} - (1/\delta)[\alpha+(\beta-\gamma_2 q)p+(\gamma_1+\gamma_2\hat{p})q+(\beta-\gamma_2 q)/\delta] \quad (16)$$

and the choke price $\hat{p} = (1/\delta)\ln[(\beta\text{-}\gamma_2 q)/\delta^2\theta]$ follows immediately from the Hicksian demand derived from (16). When evaluated at the choke price $\hat{p}$, the quasi-expenditure function (16) simplifies to the expression as before,

$$\tilde{m} = -(1/\delta)[\alpha+\gamma_1 q+\beta\hat{p}] \quad (17)$$

except that (17) is expressed in terms of the choke price $\hat{p}(\theta(q,u))$ instead of $\theta(q,u)$ because it proves more convenient to change variables and solve a differential equation in $\hat{p}$ and q. The marginal passive use value is

$$\frac{d\tilde{m}}{dq} = -\gamma_1/\delta - (\beta/\delta)\frac{\partial\hat{p}}{\partial q}. \quad (18)$$

The Hicks neutrality condition (4) is the same as in Section II, though it too is expressed in terms of $\hat{p}$ instead of θ:

$$\frac{d\tilde{m}}{dq} = [\hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial q]/[1-\hat{p}\cdot\partial x(\hat{p},q,\tilde{m})/\partial m].$$

$$= \frac{\gamma_1\hat{p}}{1-\delta\hat{p}}. \quad (19)$$

Equating (18) and (19) and simplifying, the resulting ordinary differential equation in $\hat{p}$

and q is

$$-\gamma_1/\delta + (-\beta/\delta+\beta\hat{\mathrm{p}})\mathrm{d}\hat{\mathrm{p}}/\mathrm{dq} = 0,$$

and separating the variables, integrating both sides, and solving for $\hat{\mathrm{p}}$ with the quadratic formula yields

$$\hat{\mathrm{p}} = 1/\delta \ - \ \sqrt{1/\delta^2 + 2(\gamma_1/\beta\delta)\mathrm{q} - u}. \tag{20}$$

where u is the utility index scaled from the constant of integration, and the root consistent with theory is again chosen. Using (20) in (18), the expression for marginal passive use value in terms of observables is

$$\frac{\mathrm{d}\tilde{\mathrm{m}}}{\mathrm{dq}} = (\gamma_1/\delta)[-1 + 1/\sqrt{1 + 2(\gamma_1\delta/\beta)\mathrm{q} - u}\,]\,. \tag{21}$$

It can be seen in (21) that when $\gamma_1 = 0$, marginal passive use value is zero and the quasi-expenditure function exhibits weak complementarity. In contrast to the linear demand in Section II, though, this can occur when use value is positive.

The implications of hypothesis tests on γ_1 and γ_2 can be better understood by measuring the total value (difference in quasi-expenditure functions) of a change in the amenity q, as decomposed into its constituent use and passive use parts following the approach popularized by McConnell [11]. The quasi-expenditure function for the model (15) under weak neutrality is

$$\tilde{\mathrm{e}}(\mathrm{p,q},u) = \theta \mathrm{e}^{\delta p} - (1/\delta)[\alpha + (\beta\text{-}\gamma_2\mathrm{q})\mathrm{p} + (\gamma_1 + \gamma_2\hat{\mathrm{p}})\mathrm{q} + (\beta\text{-}\gamma_2\mathrm{q})/\delta] \tag{22}$$

where $\theta(\mathrm{q},u) = [(\beta - \gamma_2\mathrm{q})/\delta^2]\mathrm{e}^{-1 + \sqrt{1 + 2[(\delta\gamma_1/\beta)q - u]}}$ and $\hat{\mathrm{p}}(\mathrm{q},u)$ in (20) are known functions of the demand parameters and covariates. The total value of a change in the amenity from q_0 to q_1 is then $TV = \tilde{e}(p,q_0,u) - \tilde{e}(p,q_1,u)$, or

$$TV = -(\Delta\theta)e^{\delta p} + (\Delta q)\Big(\frac{\gamma_1+\gamma_2(\hat{p}_1\text{-}p)}{\delta} + \frac{\beta\text{-}\gamma_2}{\delta^2}\Big) + (\Delta\hat{p})\Big(\frac{\gamma_2 q_0}{\delta}\Big)$$

where $\Delta\theta \equiv \theta_1 - \theta_0$, $\Delta q \equiv q_1 - q_0$, and $\Delta\hat{p} \equiv \hat{p}_1 - \hat{p}_0$. Total value clearly depends on both parameters γ_1 and γ_2. Nonuse value (NUV) is $\tilde{m}_0 - \tilde{m}_1$, or

$$NUV = (\beta/\delta)\,\Delta\hat{p} + (\gamma_1/\delta)\,\Delta q,$$

which depends only on γ_1 since $\Delta\hat{p}$ depends only on γ_1. Use value (UV) is

$$UV = -(\Delta\theta)e^{\delta p} + (\Delta q)\Big(\frac{\gamma_2(\hat{p}_1\text{-}p)}{\delta} + \frac{\beta\text{-}\gamma_2}{\delta^2}\Big) - (\Delta\hat{p})\Big(\frac{\beta\text{-}\gamma_2 q_0}{\delta}\Big)$$

which also clearly depends on both γ_1 and γ_2. As a result, significance tests on the amenity parameters γ_1 and γ_2 have the following implications:

$\underline{\gamma_1 = 0}$: Passive use value is zero (weak complementarity case).

$\underline{\gamma_2 = 0}$: Weak neutrality model from Section II; passive use value is nonzero generally.

$\underline{\gamma_1 = 0 \text{ and } \gamma_2 = 0}$: Use and nonuse value are zero, total value is zero.

III. In Conclusion

This paper has illustrated the procedure of integrating back from a simple empirical demand specification to recover the weakly neutral quasi-expenditure function it implies. The procedure is analogous to that used by Larson [8] to recover weakly complementary quasi-expenditure functions, but the weakly neutral quasi-expenditure function does not embody the requirement that passive use value be identically zero as weak complementarity does. The weak neutrality structure preserves the essential insight of the suggestion by Mäler [10] about private and public good linkages, but does so in a way that weak complementarity is one of many possible preference structures that satisfy the asserted private-public good linkage. That is, weak complementarity is nested within appropriately formulated weakly neutral models and as such can be tested for. The implication of rejecting weak complementarity is that the quasi-expenditure function consistent with observed demand behavior also contains an element of what is commonly termed "passive use value" following the work by [11] and others. The passive use element, in and of itself, may be of less interest than being able to link demand behavior to a well-identified quasi-expenditure function that yields estimates of the total value of amenity changes from observed behavior.

In light of the NOAA panel's encouragement of the development of methods for valuing amenity changes, the strategy of making those valuations indirectly from observation of behavior warrants further consideration. All methods require unverifiable assertions about the valuation process as reflected by individuals' actions or statements, whether in the form of assumptions that people would actually behave as they say they would or that preferences for amenities are linked in specific ways to consumption of private goods. Weak neutrality of an amenity with private goods is an example of the latter strategy that has gotten some attention recently. It is closely akin to, but not identical with, weak complementarity in its exploitation of public-private good linkages to asess the value of amenities. One of the difficulties in evaluating this alternative behavioral approach is that it has not been clear how to implement the strategy empirically. This paper has begun the process of answering that question by showing how weak complementarity nests within weak neutrality for simple demand models, and finding the corresponding quasi-expenditure functions.

Footnotes

1. More broadly, this observations applies to any welfare analyses that involve changes in non-price arguments of the individual's utility function.

2. It should be noted that the comparable unobservable assumption which must be made when using contingent valuation to infer valuations is that people would in fact pay what they say they would. Apart from the natural skepticism which such a proposition raises in many observers, recent research (e.g., Neill *et al.* [12]) casts doubt on its veracity. A primary difficulty, even with carefully-designed surveys involving more-tangible market goods, appears to be the hypothetical nature of the purchase commitment being expressed.

3. This assertion is also made when weak complementarity is invoked.

4. This corresponds with how weak complementarity has has been treated in most conceptual and empirical treatments.

5. The sign of θ is determined by the second order condition. The same parameter restrictions as in [8] apply here.

6. The resulting constant term on the right side of (11), which is arbitrary, is chosen as $\ln(-\phi)$ for convenience in comparisons with results derived under weak complementarity.

7. The same conclusion results if this restriction is substituted into the weakly neutral quasi-expenditure function (13).

8. This could be tested using the value of θ_0 from the initial conditions of the demand problem, i.e., $\theta_0 = (1/\delta)[x_0 + \beta/\delta]e^{p_0}$, where $x_0 \equiv x(p_0,q_0,m_0) = \alpha + \beta p_0 + \gamma q_0 + \delta m_0$.

References

1. Arrow, K. J., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman, "Report of the NOAA Panel on Contingent Valuation," National Oceanic and Atmospheric Administration, Washington, D.C., January 1993, 41 pp.

2. Bockstael, N. E., and K. E. McConnell, "Public Goods as Characteristics of Non-Market Commodities," *Economic Journal*, September 1993, 103, 1244-1257.

3. Dickie, M., and S. Gerking, Willingness to Pay for Ozone Control: Inferences from the Demand for Medical Care," *Journal of Environmental Economics and Management*, January 1991, 21, 1-16.

4. Flores, N., "Reconsidering the Use of Hicks Neutrality to Recover Total Value", *Journal of Environmental Economics and Management*, forthcoming, May 1996.

5. Hausman, J. A., Exact Consumer's Surplus and Deadweight Loss, *American Economic Review* September 1981, 71, 662-676.

6. LaFrance, J. T., "Weak Complementarity and Other Normative Speculations," in Western Regional Research Project W-133, *Benefits and Costs In Natural Resource Planning*, ed. R. Ready, University of Kentucky, 1994.

7. LaFrance, J. T., and W. M. Hanemann, The Dual Structure of Incomplete Demand Systems," *American Journal of Agricultural Economics*, May 1989, 71, 262-274.

8. Larson, D. M., "Recovering weakly complementary preferences," *Journal of Environmental Economics and Management*, September 1991, 21, 97-108.

9. Larson, D. M., "Further Results on Willingness to Pay for Nonmarket Goods," *Journal of Environmental Economics and Management*, March 1992, 23, 101-122.

10. Mäler, K.-G., *Environmental Economics: A Theoretical Inquiry*, Baltimore: Johns Hopkins Press for Resources for the Future, 1974.

11. McConnell, K. E., "Existence and Bequest Value," in *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas* (R. Rowe and L. Chestnut, eds.), Boulder: Westview Press, 1983.

12. Neill, H. R., R. G. Cummings, P. T. Ganderton, G. W. Harrison, and T. McGuckin, "Hypothetical Surveys and Real Economic Commitments," *Land Economics*, May 1994, 70, 145-154.

13. Neill, J. R., "Another Theorem on Using Market Demands to Determine Willingness to Pay for Nontraded Goods," *Journal of Environmental Economics and Management*, June 1988, 15, 224-232.

14. Portney, P. R, "The Contingent Valuation Debate: Why Economists Should Care," *Journal of Economic Perspectives*, Fall 1994, 8, 19-43.