WESTERN REGIONAL RESEARCH PUBLICATION

W-133
Benefits and Costs Transfer in Natural Resource Planning

Ninth Interim Report

July 1996

Compiled by
Joseph A. Herriges

Department of Economics
Iowa State University
Ames, IA 50011-0221

CONTENTS

Other

INTRODUCTION

This volume contains the proceedings of the W-133 Regional Research Project's Annual Technical Meeting held on Jekyll Island in Georgia from March 13th through the 15th. The purposes of this Western Regional project include the economic valuation of environmental amenities and natural resources; the use of these values to inform public policy and decision making; 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. This past year, for example, there were nearly twenty attendees at the annual meeting from federal and state agencies, including the U.S. Bureau of Reclamation, the U.S. Forest Service, NOAA, the U.S. Fish and Wildlife Service, Texas Park and Wildlife, and New York State's Department of Environmental Conservation. 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 non-use values of natural resources for public policy analysis; and (2) to develop protocols for transferring value estimates to unstudied areas. Research conducted by W-133 participants to meet these objectives is targeted to a number of research areas, including: water-based recreation, groundwater quality, and recreational fisheries. In addition to the many case studies of amenity values and benefit 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 research areas of the W-133 project. The first two 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. The next section addresses the objective of benefits transfer. The final three sections address some of the fundamental methodological issues that run throughout all efforts to provide convincing resource values, including the areas of contingent valuation and recreation demand modeling.

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 important areas of amenity valuation which the W-133 project makes possible.

In closing, I would like to thank Donna Otto for her capable assistance in preparing this volume.

Joseph A. Herriges
Department of Economics
Iowa State University
July 1996

HOW CERTAIN ARE VISITORS OF THEIR ECONOMIC VALUES OF RIVER RECREATION: RESULTS FROM CHALLENGING RESPONDENTS' ANSWERS

John B. Loomis
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, CO 80523

Armando Gonzàlez-Cabàn
USDA Forest Service
Pacific Southwest Station
Forest Fire Lab
Riverside, CA 92507

March 1996

ABSTRACT

While the Contingent Valuation Method (CVM) elicits behavioral intentions, the Reference Operating Conditions proposed by Cummings, et al., suggest that visitor valuations should be relatively well formed. We test the robustness of visitor dichotomous choice willingness to pay (WTP) for maintaining instream flow by challenging respondent's affirmative answers. This is followed by rephrasing the per trip WTP question into an annual dichotomous choice question and reasking the question. About 10% of visitors revised their "Yes" answer to "No". Estimated WTP changed from $12.81 per trip to $11.96 per trip. Using the method of convolutions this is an insignificant difference. This finding suggests that visitor WTP responses for instream flow appear to be well formed and robust. Comparison of the CVM values to those derived from a count data travel cost model further suggests the CVM values have convergent validity.

Acknowledgements: We thank Dr. Fred Scatena and Dr. Ariel Lugo, IITF for their cooperation and approval of the study proposal. Biologist Ernie Garcia, CNF, for his help in describing the aquatic life in the river. Michael Welsh of Hagler Bailey Inc. was most helpful in coaching us through this multiple bounded Gauss program. Jeffrey Englin, University of Nevada-Reno provided several valuable suggestions for improving the travel cost model. Last, but not least, we wish to thank Hispania Research personnel for their professionalism and quality of their work. It would have been difficult, to complete the study without their assistance in data collection and coding. Of course, the opinions expressed in this report are those of the authors and do not necessarily represent those of the Forest Service, IITF or any other of the organizations named here.

I. INTRODUCTION

Many water resource trade-offs involve reducing water for recreation to augment municipal or irrigation supplies. Since public agencies often do not charge market clearing prices for access to rivers for recreation, economists use either revealed preference methods such as the travel cost method (Ward, 1983, 1987) or intended behavior methods such as contingent valuation (Duffield, Neher and Brown, 1992).

Contingent valuation method (CVM) surveys of *visitors* to a particular recreation site generally meet what Cummings, et al. (1986:104) called the Reference Operating Conditions (ROC's) necessary for accurate valuation by CVM. These conditions are: (1) Subjects must understand and be familiar with the commodity to be valued; (2) Subjects must have had prior valuation and choice experience with respect to the consumption levels of the commodity; (3) There must be little uncertainty; (4) WTP not WTA measures are elicited.

Visitors have repeated opportunities to exchange money for the good in question. That is, in making their trip decisions, they have traded money, in the form of travel costs, for access to the river for recreation. In our study, the majority of visitors took multiple trips per year, thus providing them with the opportunity to learn as they would with repeated market transactions. This suggests the first ROC is satisfied since multiple trips provide the visitor with familiarity with the site they were asked to value. There is also little uncertainty about the good being valued as they were asked to value a trip to the site they are currently visiting. Finally, our survey elicited WTP.

Unlike the situation of asking non-visiting households about their WTP for protecting a resource they have never seen before, one would expect the values of visitors to be relatively well formed. However, Gregory, et al. (1995) question whether people have specific values for even familiar natural resources such as smoke emissions, salmon habitat in their own state (Oregon) or elk. Given the Reference Operating Conditions, we would expect visitors to be fairly certain whether they would pay a given increase in trip cost to make a visit to the site under question and not easily susceptible to changes in their responses. We test this idea by challenging their responses with a follow-up question, "Are you sure you would pay this amount?". This challenge was followed by confronting them with what this increase in cost per trip would mean to their annual cost of visiting this site to give them a subsequent opportunity to revise their answer. We evaluated the number of visitors that revised their answers as well as whether the resulting estimate of mean WTP was significantly different from that calculated using only the original responses.

The travel cost method (TCM) has the advantage of being based on actual behavior and is sometimes used to assess the convergent validity of the contingent valuation method (Mitchell and Carson, 1989:204; Sellar, Stoll and Chavas, 1985). In this study we compare WTP derived from an individual observation count data specification of the travel cost method (Creel and Loomis, 1990) with that of a relatively new variant of the dichotomous choice format contingent valuation method. This CVM questioning sequence is similar to the double-bounded dichotomous choice method (Hanemann, et al., 1991) with the addition of a third step down or spike at $1 (Hanemann and Kristrom, 1994) or generalized by Welsh and Bishop (1993) as the multiple-bounded approach. In particular, we compared the confidence intervals to determine if the resulting WTP values are significantly different between TCM and the original and revised WTP from CVM. We believe the results of this paper will contribute to understanding of the relative accuracy and robustness of WTP responses to dichotomous choice CVM surveys for river recreation and the value of instream flow.

II. HYPOTHESES

Our hypothesis that visitors have well formed values suggests that when the visitor's affirmative response to a given bid value is challenged, they will not change their response. Visitors with well formed values will also have thought out the consequences of their responses. Thus, rephrasing the WTP question from a per trip cost increase to an equivalent annual cost increase (based on the respondent's reported number of trips) may not change the affirmative response. Therefore, we would expect few visitors to switch their responses from "yes" to "no". As such, mean WTP estimated from the original responses (WTP_{orig}) should equal WTP estimated using cumulative revised responses (WTP_{rev}) resulting from the two different challenges to respondent's answers. The null hypothesis is:

(1) Ho: $WTP_{orig} = WTP_{rev}$

Application of the ROC's would also suggest that CVM derived WTP should be equal to WTP estimated using TCM as they are essentially measuring the same construct (other than the difference between compensating variation and consumer surplus associated with CVM and TCM, respectively). Thus:

(2) Ho: $WTP_{orig} = WTP_{tcm}$

(3) Ho: $WTP_{rev} = WTP_{tcm}$

III. STATISTICAL METHODS

A. Multiple Bounded Dichotomous Choice CVM

Depending on the type of CVM WTP question asked, WTP can be directly calculated (e.g., open-ended WTP responses) or estimated from a statistical model. In this study, the double-bounded dichotomous choice approach (Hanemann, et al., 1991) was modified by adding a third step down at $1. Kristrom and Hanemann (1994) have proposed this in the case of the single bound dichotomous choice model, but it has not been tried before with the double- bounded model. If individuals say "no" to the first and the follow-up lower dollar amounts, they were asked whether they would pay $1. Yes responses to the initial bid were only asked whether they would pay a dollar amount higher than the initial amount. This general pattern can be viewed as a simplified version of the multiple bounded dichotomous choice approach recently suggested by Welsh and Bishop (1993).

Each respondent was asked at least two different dollar amounts and could be asked up to three, if they said NO to the first and second amounts. Our question sequence makes five possible response combinations: (a) $y_i y_u$; (b) $y_i n_u$; (c) $n_i y_l$; (d) $n_i n_l y_{\$1}$; (e) $n_i n_l n_{\$1}$, where subscript i is the initial dollar amount asked, subscript u is the upper dollar amount asked, l is the lower dollar amount asked and $1 is the lowest dollar amount asked of individuals that said no to the lower dollar bid amount.

Response patterns (b)-(d) bracket the respondent's WTP between two of the bid amounts they were asked. Because people would not choose to visit a site unless the anticipated benefits are greater than zero, the fifth response category is bracketed from below by zero. This bracketing along the real number line is illustrated below :

```
|---------------|--------------------|-----------------------|------------------|---------------------------->

0              $1                  lower                  initial           upper
                                   $ amt                  $amt              $amt
  (e) nᵢnₗn$1      (d) nᵢnₗy$1              (c) nᵢyₗ          (b)yᵢnᵤ          (a) yᵢyᵤ
```

Using a multiple bounded approach to calculate the specific dollar amount a person would pay involves estimating the probability density function only over the bracketed interval, i.e., only the bracketed interval contributes to the likelihood function. The log likelihood function is:

$$(4)\ \ln(\text{Likelihood}) = \sum_{r=1}^{n} \ln(P_{ru} - P_{rl})$$

where, P_{ru} and P_{rl} are the probabilities that respondent r would pay their upper dollar amount (u) and lower dollar amount (l), respectively. The only difficulty is dealing with response category (a) where the yes-yes response does not allow us to observe an upper bound on the individual's WTP. However, we do know, with probability =1, that the respondent's WTP is larger than the upper amount. Welsh and Bishop (1993: 339-340) use this observation to program the log likelihood function for this first response category.

For ease in computing the log likelihood function, the probability density function of WTP is often assumed to be logistically distributed. The log likelihood function is maximized with respect to the vector of parameters (**B**'s) explaining the pattern of responses observed using a Gauss program developed by Welsh and Bishop (1993). At a minimum, the parameters include the bid amount the individual was asked to pay. Additional variables may include responses to attitude questions or the respondent's demographics such as age, education, membership in environmental organizations, etc. Specifically the log likelihood function is maximized with respect to **B** as shown in equation (5):

$$(5) \qquad \frac{\partial \ln(\text{Likelihood})}{\partial \mathbf{B}} = \sum_{r=1}^{n} \frac{1}{P_{ru}-P_{rl}} * \left[\frac{\partial Pru}{\partial \mathbf{B}} - \frac{\partial Prl}{\partial \mathbf{B}} \right] = 0$$

From the **B**'s estimated in equation (5), Hanemann (1989) provides a formula to calculate the expected value of WTP if WTP must be greater than or equal to zero. The formula is:

(6) Mean WTP = $(1/B_1) * \ln(1+\exp Bo)$

where B_1 is the coefficient estimate on the bid amount and Bo is either the estimated constant (if no other independent variables are included) or the grand constant calculated as the sum of the estimated constant plus the product of the other independent variables times their respective means.

B. Testing for Statistical Differences in CVM Derived WTP

As is evident from equation (6), WTP derived from dichotomous choice CVM responses is the ratio of two random variables. As such WTP is a random variable. To estimate the variance of WTP one could adopt a bootstrapping approach and repeatedly re-estimate the logit equation, calculating WTP for each newly estimated equation (Cooper, 1994). Alternatively, Park, et al. (1991) suggest that the information necessary to calculate confidence intervals already exists in the original variance-covariance matrix. As such, one can simply make repeated draws of coefficients using the variance-covariance matrix and calculate a WTP with each draw. We used this approach to calculate the CI's on WTP from CVM. If the CI's of WTP_{orig} and WTP_{rev} do not overlap, we would reject the null hypothesis and consider the alternative that visitor values are not stable, but rather change in a statistically significant manner if challenged or the question is rephrased from a per trip basis to an annual basis. However, if the CI's do overlap, it is difficult to conclude whether to accept the null hypothesis, and if so at what alpha level (Poe, et al., 1994).

To provide a more conclusive test and one in which the alpha levels are explicit, Poe, et al., adapted the method of convolutions to test for differences in the distribution of estimated WTP's. In particular, the distribution of mean WTP is calculated using the approach of Park, et al. for both the initial and changed WTP. Then, all possible combinations (i.e., convolutions) of the two distributions are subtracted from one another. If the differences are not statistically significant, we accept the null hypothesis of equality between WTP_{orig} and WTP_{rev}. See Poe, et al. for more details on the method of convolutions.

C. Specification of Travel Cost Method

The basic travel cost demand model is specified as:

(7) $TRIPSi = \beta o - \beta_p (TCi) + \beta_2 (TRVTIMEi) + \ldots \beta nXn + \epsilon$

where: TRIPSi is the annual number of trips person i takes to the site

TCi is the round trip costs from individual i's residence to the site.

TRVTIMEi is the travel time from individual i's residence to the site (in minutes).

TCi represents the price variable in the model. TRVTIME is included to account for the separate influence of travel time on the number of trips taken. It is well known that omission of travel time will bias the TCi coefficient resulting in a downward bias in consumer surplus estimates (Cesario, 1976, Ward, 1983).

The number of trips taken is a non-negative integer, which suggest that statistical efficiency in estimation can be improved by using a count data specification (Creel and Loomis, 1990; Englin and Shonkwiler, 1995). Given that data was collected from visitors, each person has at least one trip. This results in a truncated sample. As noted in the next section on data collection, visitors were interviewed on-site. With on-site sampling individuals who visit more often are likely to be oversampled. This leads to endogenous stratification. Englin and Shonkwiler (1995:106) show that the Poisson formulation of the count data model can be easily modified to correct for both truncation and endogenous stratification.

In particular, the basic form of the Poisson model is:

(8) $\ln\lambda_i = \beta\mathbf{X}_i + \epsilon$

and the modification of Englin and Shonkwiler is to subtract one from λ_i, resulting in (9):

(9) $\ln(\lambda_i - 1) = \beta\mathbf{X}_i + \epsilon$

Either form of the Poisson model is equivalent to a semi-log model. Therefore, the consumer surplus per trip is (Creel and Loomis, 1990):

(10) $CS/TRIP = 1/\beta_p$

where β_p is the price coefficient in the travel cost demand equation.

To test hypotheses (2) and (3) we compared the 95% CI's on WTP per trip using TCM and the respective CVM estimates. If the CI's do not overlap then CVM WTP is statistically different than WTP_{tcm}. This would suggest that even though visitor values may be robust, they may not have construct validity. However, it may be the case that the estimates are statistically different, but the difference is so small as to not affect its use in policy analysis.

IV. DATA COLLECTION

A. Survey Design

The particular application is to recreation at a river in Puerto Rico. The Rio Mameyes is threatened by a proposal to reduce its flows in half, while at the same time augmenting the sewage treatment discharges into the river. Therefore, we desired to quantify the existing recreation value. Prior to formally developing the survey instrument a focus group was held in the town closest to the river and consisted only of people who recreate in the river on a regular basis. Following this focus group, a complete survey script was developed. A cadre of interviewers were trained in the proper techniques to conduct a personal interview and then the survey was pre-tested on a small sample (n=30) of visitors. During the pre-test one of the authors accompanied each one of the interviewers to ensure consistency and quality control of the interviews. During the interviews we repeatedly probed the respondent to determine if any portions of the survey or questions were confusing or unclear. Finally, the pre-test was used to refine the range of bid amounts for the dichotomous choice WTP questions.

In the economic section of the survey visitors were first asked their trip cost. They were then asked their willingness to pay higher trip costs to visit the Rio Mameyes. Specifically, they were asked if they would still visit the Rio Mameyes today, if their cost were $XX higher than they already spent on that visit. If they said NO, the dollar amount was reduced by half. If they said NO to this amount, they were asked if they would pay $1 more to visit. If they say YES to the initial amount, the dollar amount was doubled and asked if they would pay this increase in costs.

To ascertain how certain visitors were of their responses, if they said YES to this doubled amount, they were asked if they were certain that they would pay this doubled amount. If they still responded YES, the bid amount was multiplied by their annual number of visits to compute an annual increase in costs. They were then asked if they would really pay this annual amount to continue to take their current number of visits to the Rio Mameyes under current conditions.

The bid amounts were $5 per trip to $120 per trip at the high end. These initial or starting bid amounts were based on responses to discussion in the focus groups and pre-testing of the survey questionnaire.

B. Sampling

Recreation users were sampled at two locations, at the mouth of the river and at a site that will be referred to as the restaurant site as it is next to a closed restaurant. Surveys at the restaurant site were performed on half the weekends in July and August as well as two holidays and weekdays for a total of 12 days during 1995. Surveys at the mouth of the river were conducted on half the weekends in July as well as two holidays and two weekdays for a total of nine days during 1995. Recreation users were interviewed on-site. One person from every group present at the site during survey period (10am to 5pm) was interviewed. Visitors were screened for minimum age of 16 (i.e., driving age so they could make their own trip decisions). In addition, we did not interview visitors who had been previously interviewed at the recreation site.

V. RESULTS

A. Response Rate

A total of 274 recreation users were contacted and 200 interviewed for a response rate of 73 percent (200 interviewed/274 contacted). Common activities were wading, swimming, diving and picnicking.

B. Pattern of Responses With and Without Challenging Visitors

As shown in the middle of Table 1, 152 of the 191 respondents (76%) said YES to the initial bid amount (Bi). Of this group, 91 of the 152 (60%) said Yes to the higher bid amount (Bu). All 199 of these responses were used to estimate the initial multiple bounded logit equation in the Original Response column in Table 2.

The group of Y_iY_u responses were then challenged by the interviewer. The first challenge was "Would you really pay this bid amount?", to which 85 out of the 91 (93%) indicated they would (Table 1). This group was further challenged by confronting them with what paying this bid on each of the trips they take over the year implied on an annual basis (i.e., Bu * TRIPSi). About 96% (82/85) responded they would pay this amount each year to have access to the river for recreation under existing conditions. Across both challenges nine respondents changed their Yes answers to No.

To estimate WTP_{rev} we recoded the original Yes responses at Bu to No for those nine individuals who when challenged, indicated they would not really pay Bu. Using the recoded data we re-estimated the multiple bounded logit equation and these estimates are in the Response after Challenge column of Table 2.

As shown in Table 2, the coefficients on age, travel time and bid were all statistically significant. WTP was calculated using equation 6 and is reported in Table 4. The 95% CI's were calculated using the method of Park, et al., and are also reported in Table 4.

C. Travel Cost Method Results

Table 3 presents the results from the Poisson regression modified to incorporate truncation and endogenous stratification. Trip costs, travel time and age were statistically significant at the .01 level. Using the formula for WTP in equation (10), mean WTP from TCM is $73.76 per group trip or $15.43 per person per trip. The confidence interval on the per person per trip WTP is $11 to $26 (Table 4).

D. Results of Hypothesis Tests

As shown in Table 4, the 95% CI's for WTP_{orig} and WTP_{rev} have a substantial overlap, with the lower bound of WTP_{orig} overlapping the mean of WTP_{rev}. Since the 95% CI's for WTP_{orig} and WTP_{rev} do overlap, the method of convolutions was used to verify that there was no significant difference between the two WTP estimates. The convolution included zero (meaning the differences in the distributions of mean WTP were not statistically different from zero) and the alpha significance level was .368. Thus, visitors' WTP values seem quite robust and well formed. Comparing the mean WTP and CI's derived from TCM suggests that for visitors, the TCM derived WTP are not statistically different than WTP from CVM since both the CVM CI's are contained within the TCM CI. In this case TCM and CVM derived estimates of WTP have convergent validity, but the revealed behavior approach yields a WTP <u>greater</u> (but not statistically greater) than intended behavior.

VI. CONCLUSIONS

Willingness to pay of visitors appears to be well formed and robust. When individuals were challenged regarding their affirmative answer to their bid amount, only 6 out of 91, revised their amount downward. When the per trip WTP question was reformulated as an annual WTP, only 3 of 85 respondents switched from an affirmative response. This stability of visitor responses may be due to several factors. Optimistically, we believe it is because they do have a good idea of what they would pay. This conclusion rests on the fact that asking visitors' their WTP meets several of the Reference Operating Conditions suggested by Cummings, et al. In particular, a majority of our visitors took multiple trips per year to the site. Thus, they had ample experience with the good in question and have repeatedly traded travel cost for access to the site. Further, the WTP values derived from the CVM responses were not statistically different than WTP from the revealed preference TCM.

A less optimistic explanation is that once visitors committed to an affirmative response they were reluctant to change if for fear of appearing to have lied to the interviewer. However, this explanation seem inconsistent with the slight underestimate of WTP by CVM relative to TCM. The CVM underestimate suggests that visitors would have paid at least what they stated to the interviewer. It would have been informative to follow these affirmative responses up with an actual cash transaction on-site to evaluate the criterion validity of such firm responses among visitors. This last task must await further research.

REFERENCES

Cesario, Frank. 1976. Value of Time in Recreation Benefit Studies. *Land Economics* 52(1):32-41.

Cooper, Joseph. 1994. A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys, *Land Economics* 70(1):111-122.

Creel, Michael and John Loomis. 1990. Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California. *American Journal of Agricultural Economics* 71:434-441.

Cummings, Ronald, David Brookshire and William Schulze. 1986. *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method.* Rowman and Allenheld, Totowa, NJ.

Duffield, John, Christopher Neher and Thomas Brown. 1992. Recreation Benefits of Instream Flow: Application to Montana's Big Hole and Bitterroot Rivers. *Water Resources Research* 28(9):2169-2181.

Gregory, Robin, Sarah Lichtenstein, Thomas Brown, George Peterson and Paul Slovic. 1995. How Precise Are Monetary Representations of Environmental Improvements. *Land Economics* 71(4):462-473.

Englin, Jeffrey and J.S. Shonkwiler. 1995. Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation. *Review of Economics and Statistics* 77(1):105-112.

Hanemann, Michael. 1989. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses Reply. *American Journal of Agricultural Economics*: 71:1057-1061.

Hanemann, Michael, John Loomis and Barbara Kanninen. 1991. Statistical Efficiency of Double Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics* 73:1255-1263.

Hanemann, Michael and Bengt Kristrom. 1994. Preference Uncertainty, Optimal Designs and Spikes in Environmental Economics, edited by R. Brannlund, B. Kristrom, K. Lofgren and L. Mattsson. Dept. of Forest Economics, Swedish Univ. of Agric. Sciences, Umea, Sweden.

Mitchell, Robert and Richard Carson. 1989. Using Surveys to Value Public Goods: the Contingent Valuation Method. Resources for the Future, Washington DC.

Park, Timothy, John Loomis and Michael Creel. 1991. Confidence Intervals for Evaluating Benefits from Dichotomous Choice Contingent Valuation Studies. *Land Economics* 67:64-73.

Poe, Gregory, Eric Severance-Loss in, and Michael Welsh. 1994. Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach. *American Journal of Agricultural Economics* 76(4): 904-915.

Sellar, Christine, John Stoll and Jean-Paul Chavas. 1985. Validation of Empirical Measures of Welfare Change: A Comparison of Non-market Valuation Methods. *Land Economics* 61(2):156-175.

Ward, Frank. 1983. Measuring the Cost of Time in Recreation Demand Analysis: Comment. *American Journal of Agricultural Economics* 65:167-168.

Ward, Frank. 1987. Economics of Water Allocation to Instream Uses in a Fully Appropriated River Basin: Evidence from a New Mexico Wild River. *Water Resources Research* 23(2): 381-392.

Welsh, Michael and Richard Bishop. 1993. Multiple Bounded Discrete Choice Models. in Bergstrom, J., Compiler. W-133 Benefits and Costs in Natural Resource Planning, Sixth Interim Report. Dept. of Agricultural and Applied Economics, University of Georgia, Athens, GA.

Table 1. Response Pattern

Original Response Pattern (Unchallenged) (n=199)					
	NO				
Initial Bid (Bi)	47				
		Yes	No		
Lower Bid (Bl)		29	18		
				Yes	No
One Dollar (B1$)				17	1
	YES				
Initial Bid (Bi)	152				
		Yes	No		
Higher Bid (Bu)		91	61		

Changes in YES-Yes Response Pattern with Challenges (n=91)				
	No	Yes		
Would you really pay Bu per trip?	6	85		
Would you really pay Bu * Trips each year?			Yes	No
			82	3

Table 2. Multiple Bounded Logit Equations

	Original Responses		Responses After Challenge	
Var	Coefficient	T-Stat	Coefficient	T-Stat
Constant	1.1225	2.469	1.2406	2.771
Age	0.0318	2.811	.0266	2.440
Trvltime	0.0099	2.211	.0110	2.446
BID($)	-0.0451	-12.27	-.0479	-12.632

	Original Responses	Responses After Challenge
-2*Log Likelihood:	530.19	545.956
Wald Statistic:	151.349	159.966
Probability of a larger Wald Stat:	0.00	0.00
Degrees of freedom:	196	196

where: Age is age of respondent in years; Trvltime is one-way travel time of respondents to the river from their home.

Table 3. Poisson TCM Results

Variable	Coefficient	T-stat	Mean
Constant	3.047	45.35	
TC	-.0135	-4.68	9.42
Trvltime	-.0257	-23.75	44.67
Age	.0112	7.72	35.55

Log-Likelihood: -2347
Restricted Log-Likelihood: -2806
Likelihood Ratio Statistic: 918

Table 4.
Comparison of Original and Revised WTP per Trip from CVM with WTP from TCM

	$CVM\text{-}WTP_{orig}$	$CVM\text{-}WTP_{rev}$	TCM-WTP
Lower 95% CI	\$11.52	\$10.78	\$11.01
Mean	**\$12.81**	**\$11.96**	**\$15.43**
Upper 95% CI	\$14.23	\$13.22	\$26.55

A METHOD FOR NONRESPONSE CORRECTION IN THE ANALYSIS OF MAIL SURVEY DATA

by

Trudy Ann Cameron*
Department of Economics
University of California, Los Angeles, CA

W. Douglass Shaw
Department of Applied Economics and Statistics
University of Nevada, Reno, NV

Shannon R. Ragland and Sally Keefe
Hagler Bailly, Inc., Boulder, CO

John M. (Mac) Callaway
UNEP Collaborating Centre on Energy and the Environment
Risoe National Laboratory, Roskilde, Denmark

ABSTRACT

Even when researchers using mail surveys go to great lengths to maximize response rates, there will always be some portion of the intended sample that does not respond, either at all, or sufficiently completely to allow inclusion in the estimating sample. This paper examines nonresponse and its apparent consequences for a survey of water-based recreational participation conducted in the Northwest US. We describe how zip codes can be used in combination with special software to determine distances from each address in our intended sample of 7034 households to each of the recreational sites featured in our survey. Zip codes also allow us to merge our intended sample with 1990 Census data aggregated to the zip code level. We demonstrate how to model the survey response/nonresponse decision explicitly, and show that statistical corrections for nonresponse can have a potentially important effect on the apparent inferences from our models. We strongly advocate, based on these results, that any researcher using a mail survey can and should explore analogous response/nonresponse models and corrections before making any claims as to the robustness of empirical results to non-random sample selection.

* This data set employed in this paper was collected as part of a study managed by Hagler Bailly Consulting, Inc., under contract with the U.S. Army Corps of Engineers (contract # TCN 93357).

1. INTRODUCTION

The purpose of this paper is to demonstrate the opportunities for, and utility of, explicit modelling of survey response/nonresponse. A good understanding of the relationship between survey response propensities and observable relationships among the subsample of respondents can help inform researchers and policy makers about the likely nature of nonresponse biases. Mail surveys, in particular, have long been a popular method for gathering research information. They continue to be employed in a wide variety of disciplines where household decisions, preferences, or behaviors need to be quantified. A perennial concern with mail surveys is the maximization of response rates (Dillman, 1978). However, even with aggressive campaigns of follow-up reminder postcards, nominal payments to respondents, and replacement mailings, there almost always remains a persistent nonresponse group.

The issues discussed in this paper are relevant to a very broad community of investigators who rely on data gathered from mail surveys, but the discussion here will be cast in terms of an example where a mail survey has been used to collect demand information concerning non-market environmental goods (economic models used with these types of data have included travel cost models, random utility models, and contingent valuation or behavior models).

Section 2 of this paper reviews a selection of findings concerning survey response/nonresponse that have appeared in the broader marketing and social science literature, as well as a small number of studies focussing on this issue within the boundaries of environmental economics. Section 3 covers the manner in which these earlier insights are reflected in the modelling of nonresponse in our specific illustrative environmental valuation context. Section 4 outlines a rudimentary model of water-based recreational trip-taking, to be estimated using the sample of respondents to a mail survey--with and without corrections for selectivity. Section 5 formally lays out the log-likelihood function that would ideally be maximized in order to simultaneously identify the parameters of both the response/nonresponse model and the trips model, plus the correlation between the error terms in these two submodels. We also explain how best to proceed if this function proves impossible to optimize for a given data set, as is the case here.

Section 6 provides a discussion of the empirical findings in our specific example, focussing upon some possible consequences of failure to correct for nonresponse. Section 7 concludes with comments on the apparent implications of our simple illustration for more-formal survey-based estimates of demand and welfare in other contexts.

2. REVIEW OF THE RELEVANT LITERATURE ON RESPONSE/NONRESPONSE

Much of the general survey research literature on nonresponse has been devoted to studies of the relative effectiveness of different techniques that might be used during survey design and administration to maximize response rates. Fox et al. (1988) describe a meta-analysis of some of these techniques as they apply to mail survey responses. The classes of factors they consider are under the control of the researcher and include several related to the content of the cover letter, the amount and type of incentives offered, the form of respondent contact and follow-ups, the type of postage used for outgoing and return mail, and the topic, length, color, complexity and format of the questionnaire itself.

Among environmental economists, Dillman (1978) has been the standard handbook on design methods to maximize response for mail and telephone surveys.[1] Maximizing response rates is extremely important, but we are concerned here with the task of correcting demand and welfare models for any non-response that remains. To do this, we must acknowledge that nonresponse results from the decision-making process of survey recipients. Heterogeneity in sociodemographic characteristics across the intended sample may account for a significant portion of the systematic differences in response rates.

McDaniel et al. (1987) provide an excellent summary of survey research that focusses on the demographic characteristics of non-respondents. They also cite research that investigates psychographic or behavioral differences between respondents and non-respondents. Paraphrasing their summary, non-respondents to surveys tend to be less educated, of lower socioeconomic class, white, of foreign-born parentage, older, married, residents in urban areas, and living in the Northeast U.S. Psychologically and behaviorally, nonrespondents also tend to be less emotionally stable, less effective as employees, less gregarious, lower in sense-of-leadership, less widely read, less proficient in writing ability, low on order and dependency but high on aggression, dominance, autonomy and intraception, and less responsible, less tolerant, and less intellectual in personality characteristics.

McDaniel et al. (1987) make the point that this assortment of apparent tendencies from individual studies of nonresponse may not be generalizable to other studies. Their results do strongly support the common contention that the "salience" of the survey topic to the survey recipient can have a substantial bearing on the probability that the survey will be completed and returned. This conforms with an earlier meta-analysis by Heberlein and Baumgartner (1978).

Regional differences in populations may have an effect on response rates. (See, for example, Jobber and Saunders (1988), Jobber et al. (1991), Ayal and Hornick, 1986, and evidence of Canadian-U.S. differences in Goyder (1985).)

Previous Solutions

Mitchell and Carson (1989, pp. 267-282) review the problem of non-response as it affects contingent valuation surveys. They review econometric methods for sample selection bias correction but conclude that "Unfortunately, these methods may be of limited use in contingent valuation studies when little or no information is available on factors affecting the probability of responding to the survey. ... We know of no CV study that has attempted to use these techniques to correct for sample selection bias."

The most common strategy for addressing non-response in environmental valuation surveys is to provide marginal means for a limited set of sociodemographic variables (e.g. income, age), calculated for the respondent sample and for the population it is intended to represent. If these means are similar, little more is said. The problem is that even though respondents and non-respondents may appear similar on a selection of observable sociodemographic attributes, there may be important unobservable forms of heterogeneity that affect both response propensities and demand for the environmental good. For example, respondents to a recreational fishing survey may tend to be more-avid anglers than nonrespondents, and avidity may not be measured.

Most researchers have treated survey non-response as a problem that has no easy solution. Whitehead (1991) asserts that correction for self-selection bias requires information about nonrespondents, obtained either through screener surveys or follow-up surveys. Edwards and Anderson (1987) emphasize that "from a practical standpoint the test for selection bias resulting from nonrespondents' self-censorship" requires that one "interview a high percentage of nonrespondents." They note that "This need presents a substantial, technical challenge for contingent valuation studies." In contrast, the present paper offers a tractable general strategy for modelling and correcting for nonresponse.

There are only a very few cases in the existing literature on environmental valuation via survey-based methods where researchers have attempted to control for nonresponse bias. Edwards and Anderson (1987) limit their empirical analysis to cases of questionnaire item nonresponse, rather than complete unit nonresponse. In particular, they find that omission of observations due to protest bids or zero willingness-to-pay does not appear to produce any additional selectivity bias. Aggressive nonresponse conversion efforts allowed them to achieve an eventually very high response rate, but no data were available on non-respondents who remained.

In two other cases, the task of nonresponse evaluation has been facilitated because the researchers have access to supplementary databases where other *individual-specific* information can be linked to each targeted potential respondent. (See Whitehead et al. (1993), Englin et al. (1996), and Fisher (1996).

This paper differs from previous and concurrent efforts to explore nonresponse bias in the environmental economics literature in that it illustrates a technique that can be applied with any mail survey conducted in the United States.[2]

3. MODELLING PROPENSITY TO RESPOND TO OUR SPECIFIC MAIL SURVEY

The mail survey data we will use for our illustration comes from a four-version survey of water-based recreational participation (at lakes, reservoirs, and rivers) within the Columbia River system in the U.S. states of Washington, Oregon, Idaho and Western Montana, plus the southern portions of the Canadian provinces of British Columbia and Alberta. The larger study is discussed in detail in Callaway et al. (1995) and portions are also summarized in Cameron et al. (1996).

A copy of one of the four different versions of our survey instrument was sent to each of 7034 addresses. We are interested, first of all, in modelling the propensity for each copy of the survey instrument to be returned. Of these targeted households, 2513 returned surveys that were sufficiently complete for their data to be included in our demand analyses, for an overall response rate of 35.73%.[3]

Another task we face is to control for the fundamentally different expected response rates in different sampling strata. It is not always feasible to rely upon a strictly representative sample from the general population in modeling the demand for environmental amenities. In many cases, it is expedient to combine a basic general population samples with other convenience samples. In our illustration, for example, we over-sample people who live in close proximity to the environmental good to be valued. We also include people who are known to be users of the resources in question, by intercepting them on site. We also incorporate a subsample of potential foreign users, drawn from major urban areas of the nearest cross-border regions. Dummy variables identifying our four auxiliary strata groups are also included in our response/nonresponse model.

For this survey, Dillman's prescriptions were followed as closely as possible, in order to maximize overall response rates. There are some key features of the survey design, however. First, each of the four different versions focusses on a separate geographic region. The visual aids (photographs, both actual and computer-modified) that accompanied each questionnaire were different. While effort was made to make the written portions of the survey as comparable as possible, their different regional emphases may have resulted in differing appeal or salience to different types of respondents. In our model, response decisions are allowed to vary systematically according to survey version.

The zip code information is the key to generating variables that may go part way towards capturing the salience of the survey topic as well as the demographic or socioeconomic characteristics of the potential respondent's neighborhood. In designing our response/nonresponse model, we need to keep in mind that for a mail survey in the US, often nothing is likely to be known about each member of the intended sample beyond their mailing address (including zip code) and what version of the survey they were sent. Even if a survey research firm protects respondent confidentiality by redacting the name and street address of each respondent, it is possible that zip codes can be retained, since they rarely would allow unique identification of any respondent.

We thus attempt to capture salience of the survey topic to each respondent using an array of proxy variables. Even from the cover of the survey--the first thing a respondent would see--our survey topic is easily construed to be water-based recreation. Each recipient's actual and potential experience with water based recreation could be expected to influence response propensity. One way to attempt to capture this potential experience is by using distances between the recipient's home and each of the bodies of water featured in their particular version of the questionnaire. First, we employ the individual distances between the recipient's origin zip code and each of the three or four specific "Federal Projects" along the Columbia River system which are singled out on his or her particular version of the questionnaire. These projects are either reservoirs behind hydroelectric dams or run-of-river stretches below these dams. The identities of these waters differs by survey version, so we interact each of these distances with dummy variables for each version. We have also calculated the distances from each respondent's zip code to each of the five nearest "other" fresh waters in the target region that were not included among the set of Columbia River projects. The average of these five distances is used to represent the accessibility of other nearby water recreation opportunities. This average is also interacted with survey version dummy variables, since the set from which these "other" waters is drawn differs across survey versions.[4]

Our distance data were calculated using the ZIPFIP computer program (Hellerstein et al., 1993). Given origin and destination zip codes, this software allows the user to generate approximate road distances between the centers of these zip codes. These distances are constructed from "great circle" distances, modified by a factor (unique to each state) that converts these distances into average road distances in a manner that controls for differing densities of roadways in each state. We identified the zip codes containing (or nearest to) each of the water bodies in our study region. In combination with the zip code for each potential respondent's address, we merge these distances with the response/nonresponse data.

The zip codes for the intended sample are also the key that allows us to merge the data set with a wide array of variables from the 1990 Census (available from the STF3 data tapes). All variables are descriptors of the zip code area, rather than the individual, but to the extent that the geographical areas covered by zip codes are relatively homogeneous, some portion of the heterogeneity in these characteristics across survey recipients can be captured by these aggregate data. The Census data provide zip code populations as well as counts of persons in each of a variety of categories. Appendix Table A-1 gives the Census-based variables we have considered, with details concerning how these were calculated from the constituent variables (using the variable names from the Census tapes).

Census variables other than these ones may be important predictors of response/nonresponse in other applications.

A portion of our intended sample also consisted of Canadians. The general nonresponse research cited in the previous section certainly suggests that response rates should be allowed to vary systematically by country. If there is enough heterogeneity between different subregions of the Northwest U.S.--along the dimensions suggested for Canadians--there is good reason to allow for possible regional variation in response rates there based on similar arguments. Within the U.S., however, rather than using regional jurisdictional dummy variables such as states or counties, we elect to rely directly upon sociodemographic variations. These are the factors that such dummy variables would presumably be capturing.

It is also potentially important to control for any variations in survey format across the sample. Many environmental valuation surveys rely upon contingent scenarios (either contingent valuation or contingent behavior). These scenarios differ across versions of the survey instrument and these differences could conceivably influence response rates. In the wake of the debate about "embedding," other types of environmental valuation surveys have been designed to assess the effectiveness of different amounts of context for the valuation exercise. The level of descriptive detail for each scenario involved may differ a little or a lot across the individuals who make up the intended sample. Alternately, the nature or scope of the good to be valued may differ across survey version. These variations in the survey instrument may themselves lead to differential response rates, and this possibility has not generally been pursued in the literature on non-market resource valuation.

4. MODELLING WATER-BASED RECREATIONAL TRIPS

The first stage in our analysis (Stage A) is a discrete choice (probit) model of the response/nonresponse decisions among the intended sample. The second stage, conditional on response, is a model for the number of water-based recreational trips taken by respondents. We demonstrate the empirical importance of controlling for nonresponse in two different types of trip demand specifications (Stage B1 and B2). In Stage B1, we model the individual's total demand for trips to *any* water in the region featured on the questionnaire, regardless of specific destination. (Since the set of relevant waters varies across versions, we estimate these second-stage models separately for each version.) We construct a rough proxy for "accessibility" of water recreation opportunities by calculating the average distance from the respondent's home zip code to the nearest five waters (be they federal projects or "other" waters). Ex ante, one would expect that the less accessible these recreation opportunities--i.e. the greater this average distance--the fewer trips an individual will take.

In an alternative specification, Stage B2, the dependent variable is trips to *one* particular water in the choice set on a particular survey version. This allows us to attain crude estimates of the apparent own- and cross-price effects.[5] We illustrate the potential consequences of nonresponse bias in these disaggregate specifications using one individual water from each of two versions of the survey.

Our analysis is intended to illustrate the that nonresponse selectivity effects are potentially a very important consideration, regardless of the demand specification employed.

5. FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATION AND A COMPROMISE

1347 of our 2513 respondents took positive numbers of trips and these trip-takers averaged 12.4 trips apiece, with a standard deviation of 16.4 trips. A continuous distribution is assumed to be an adequate approximation to the conditional distribution of trip-taking propensities, suggesting a Tobit-type model for trips, in order to accommodate the sizeable observed frequency of zero trips.

Let $y_i^* = x_i'\beta + \epsilon_i$ be the latent propensity to return a completed response to the questionnaire that was mailed to household i. Since y_i^* is unobservable, the response/nonresponse outcome associated with each mailing is evaluated in terms of the associated observable variable $y_i = 1$ if the questionnaire is returned completed and $y_i = 0$ if the questionnaire is either not returned, or is returned insufficiently complete to be included in the analysis.[6] The vector of variables x_i includes Census zip code characteristics, variables that capture the differences among survey versions, the different sample strata, and the distance variables that partially proxy for the probable salience of the survey topic to targeted households.

Let $q_i^* = z_i'\gamma + \nu_i$ be each respondent's propensity to take water-based recreation trips to water bodies in the geographical area stipulated in each version of the questionnaire. If $q_i^* > 0$, then observed water recreation trips $q_i = q_i^*$. If we have $q_i^* \leq 0$, then observed trips will be $q_i = 0$.

We assume y_i^* is distributed N(0,1), with variance normalized to unity because the binary nature of y_i will not allow us to discern the scale of y_i^*. Let q_i^* be distributed $N(0,\sigma^2)$, since the observable portion of q_i^* does allow the scale to be identified. We wish to jointly model both the individual's decision about whether to respond to the questionnaire, and, conditional on response, the number of trips taken. This is a Tobit model with a sample selection correction, ideally estimated by Full Information Maximum Likelihood (FIML). For a description, see Greene (1995, p. 624), summarized here in Appendix 1.

A full-information maximum likelihood (FIML) Tobit model with sample selection is available in the LIMDEP econometric software package as a one-line command. We have also experimented extensively with programming and estimating the FIML log-likelihood directly using the GQOPT general nonlinear function optimization package (Goldfeld and Quandt, 1995). While, in principle, this log-likelihood is valid, it is notoriously difficult to optimize, even starting from the consistent parameter estimates produced by two-stage models. We have found that even very trivial specifications, with only one or two regressors, fail to converge successfully.

Given the frustrations of FIML estimation, we opt to rely upon consistent (though not efficient) two-stage estimates in the tradition of Heckman (1979). Note that the variance-covariance matrix obtained for the second-stage Tobit model is not valid. We correct it using the method of Murphy and Topel (1985). See Appendix 2.

Another minor inconvenience is that a consistent estimate of the error variance for the Tobit latent variable is necessary before a point estimate of the predicted number of trips can be recovered from the second-stage Tobit model. The necessary calculations require the point estimates of β from the first-stage probit model. (See Appendix 3.)

The potential consequences of ignoring the problem of nonresponse to mail surveys for environmental valuation are too important to forestall examination of the issue until FIML estimation of models in this genre can be rendered generally tractable. Thus we proceed below with corrected two-stage estimation methods.

6. RESULTS

The list in Appendix Table A-1 is an inventory of all of the Census variables which were examined in preliminary models. Only those variables that were persistently statistically significant determinants of response rates across a variety of exploratory specifications are included in the models to follow. In other applications, different variables may prove important.

Table 1 provides descriptive statistics for the entire intended sample of 7034 addresses. This is the universe of addresses to which questionnaires were mailed. The variables (described briefly in the body or footnotes to the table) either describe the type of subsample or are obtained by utilizing zip codes to calculate distances or to merge with the available Census data.

Table 2 gives the results for a pooled-data probit model that uses all 7034 addresses in the intended sample and attempts to explain provision of a usable response as a function of everything known about the zip code of the target household, the type of subsample it belonged to, and the version of the survey it received. We find that the "Phase 1" and "known-user" samples were statistically significantly more likely to respond. For known users, the subject matter of the questionnaire is undeniably salient, so this is not surprising. In contrast to expectations, the Canadian sample (population 5) does not appear to be statistically less likely to respond. However, this subsample is very small, and while comparable distance data were calculated by hand, no Census data were available for this sample, so the HAV-CENSUS indicator variable is highly correlated with membership in the Canadian subsample. Thus our finding may not be conclusive.

Version 3 of the questionnaire appears to have produced systematically larger response rates. Distances to three of the specific waters described in the questionnaire significantly influenced response rates for recipients of version 2 of the questionnaire. The effect was negative for one water, and positive for two others.[7] Distances to the nearest "other" waters appears to matter only for versions 1 and 3 of the questionnaire.

Among the Census variables examined, the most robustly individually statistically significant variables were those intended to capture language isolation, proportion on public assistance income, proportion urban, and proportion on social security income. Language isolation decreases response probabilities, as does a greater neighborhood prevalence of public assistance income. Recipients in more urbanized areas are more likely to respond, as are those from areas with higher level of employment in agriculture, fisheries or forestry, although the last is not statistically significant. Response propensity is also significantly higher, the greater the proportion of the neighborhood receiving social security income.[8]

Table 3 gives descriptive statistics for the sample of usable responses from each of two versions of the survey. For these subsamples, we have actual respondent-specific individual sociodemographic information, which certainly involves less measurement error than the zip code proportions used for all observations in the first stage response/nonresponse model. The individual "home zip code to each water" distances are the same values that were computed for the entire intended sample of 7034, so they are reused to calculate the accessibility variables used here. A Heckman-type two-stage method is used below, involving the additional explanatory variable constructed from the first stage: IMR. This is the standard inverse Mill's ratio for sample selectivity correction.

Table 4 contains results for the second stage Tobit regression models, for total trips to any water in a given region. We report the nonresponse-corrected and uncorrected estimates only for versions 1 and 3 of our survey, as nonresponse bias appears not to be statistically significant for versions 2 and 4 under this type of specification. Focussing first on the corrected models (the first column of results in each pair), note that there are many individually statistically significant parameter estimates.

Recall that the distances in this model have been simplified to measure an index of average distances to the nearest five waters in the region featured on the questionnaire (be it a focus water for that version, or any other water). The coefficient on this distance index is very strongly statistically significant. The incremental effect of variations in income is positive in both cases, but not significant. Age has a negative effect on trips in both versions. Not surprisingly, the (potentially somewhat endogenous) variables indicating possession of a fishing license or a boat are strongly significantly correlated with the latent number of trips underlying this type of Tobit specification.[9]

The coefficient on the IMR variable provides insights into the nonresponse bias. In models with an OLS second stage, the coefficient on the IMR is typically interpreted as the product of the correlation between the error in the response model and the (latent) trips model (ρ), and the error standard deviation in the latent trips model (σ). If ρ is zero, $\rho\sigma$ will be zero. The coefficient on IMR is statistically significant for survey versions 1 and 3, and in each case, is negative and rather large. This suggests a substantial negative correlation between the response/nonresponse decision and the trip-taking behavior for these two subsamples.[10]

The second column in each pair in Table 4 shows the results of an analogous model estimated without benefit of control for non-random nonresponse via the IMR term. The coefficient on the IMR term reveals that failure to control for nonresponse will lead to underestimates of the number of trips. For all versions of the survey where the selectivity effects are significant, our models suggest that unobserved factors which make targeted households *less likely* to respond to our questionnaire than our response model predicts also make them *more likely* to take water-based recreational trips than our trips models would predict. While it is sometimes difficult to label these unobserved factors, a reasonable speculation would be that these factors include tastes for outdoor activity, general levels of physical health and energy, and/or family composition, among others. One interpretation is that people who are busy engaging in activities related to outdoor freshwater-based recreation are too busy to waste time responding to our questionnaires. Alternatively, infrequent participants and non-users of these waters may have found it much easier to fill in our questionnaire. Rather than remembering numbers of trips to each site in different months, these people would simply have to fill in a lot of zeros.[11] It is worth noting that the finding of a negative error correlation is at odds with the common assumption that households with higher participation in an activity should be more likely to respond to surveys about that activity.

Setting to zero the correlation between the errors in the response/nonresponse model and the trips model is equivalent to eliminating the inverse Mill's ratio term from the calculation of fitted trips in the second-stage Tobit model. As expected, given the strongly significant negative coefficients on this term for versions 1 and 3, removal of sample selectivity results in the implication that a truly random sample from the population would have predicted much higher numbers of trips on average, and therefore greater aggregate utilization of the resource than implied by any demand model ignoring the selectivity problem. The differences are summarized in the last row of Table 4.

We should also note that failure to control for nonresponse can also distort the coefficient on the distance variable. For version 1, the corrected coefficient is -39, while the uncorrected one is -43, an overestimate of responsiveness. For version 3, the corrected estimate is -26, whereas the uncorrected one is -20, an underestimate of responsiveness. Clearly there is no generalizable bias. The AVG-DIST variable plays the role that the price variable would play in more-sophisticated consumer demand specifications. The estimated price coefficient is typically a key ingredient in consumer's surplus calculations. This offers some evidence that any eventual welfare estimates could potentially be distorted in a nontrivial fashion by failure to account for nonresponse. The distance (price) variable is probably most affected because distances are embodied in the IMR nonresponse correction terms. Few of the other slope parameters in this model appear to be seriously distorted by failure to correct for selectivity.

In our second type of specification, we explore a more-standard demand specification using our illustrative sample. These models concern demand for trips to a single site, and include travel costs for that site and other individual sites mentioned in that version (as well as "other" waters). Due in part to the smaller numbers of trips to individual waters, the two-stage Tobit specification did not converge for all individual waters.[12] For version 1, we illustrate with a demand model for Lake Koocanusa (W3) and for version 4, we provide estimates for demand for Lake Pend Oreille (W4).

We can now interpret the round-trip travel cost variables as prices in these simple demand models. The own-water price effects are negative and significant in the corrected specifications (with statistically significant IMR terms), suggesting downward-sloping demand curves. There is also evidence from version 1 that Hungry Horse Reservoir (W1) is viewed by recreationists as a substitute for Lake Koocanusa. Hungry Horse is often an overflow recreation site for Koocanusa. For version 4, it appears that Lake Roosevelt (W1) is a substitute for Lake Pend Oreille. Camping is the most important activity at Lake Roosevelt, as at Pend Oreille, and the large population centers in the region are located between these two sites, so it is reasonable that they might be viewed as substitutes. In no case is the average distance to the nearest five "other" waters influential.

The effect of holding a fishing license seems to matter for trips to Lake Koocanusa, but not for Lake Pend Oreille. Fishing is the most popular activity at Koocanusa, and most fishing there is done from boats. Boat ownership is significant for Koocanusa, but is not important in explaining trips to Pend Oreille. At Pend Oreille, camping and picnicking are the most important activities, so fishing licenses and boat ownership may well have not much of an effect on demand for this water.

Potential distortions to the demand relationship because of nonresponse bias remain our primary consideration in these examples, however. In both of the corrected models in Table 5, the inverse Mills ratio term, IMR, is strongly statistically different from zero and negative. This again suggests that unobserved heterogeneity that makes recipients more likely to respond to the survey also makes them less likely to take trips to each of these waters.

Comparison of the corrected and uncorrected models in Table 5 reveals the implications of failing to control for nonresponse. The magnitude of the own-price effect for Lake Koocanusa (in Version 1) is distorted slightly upwards, while that for Pend Oreille (in Version 4), is distorted substantially upwards (from -40 to -57). If the negative slope for quantity as a function of price is too great, the demand curve as usually depicted will be too flat. Welfare estimates such as consumer surplus, based on projection of the estimated demand curve up to the choke price, will imply too low a choke price if selectivity is not recognized in the estimation process. Resulting consumer's surplus estimates will then be too small (at least in these specifications--recall that the biases were mixed in Table 4 for total trips).

The apparent substitutability of Koocanusa for Hungry Horse and Pend Oreille for Roosevelt is also exaggerated if selectivity effects are ignored. Existence of satisfactory substitutes lessens the impact of compromises in the quality or availability of a particular water. The false impression that good substitutes exist could lead to undervaluing of social losses due to damage or reduced access to any of these waters.

7. CONCLUSIONS AND SUGGESTIONS FOR SUBSEQUENT MAIL-SURVEY-BASED RESEARCH

We have demonstrated that non-random nonresponse to a mail survey has the potential to cause substantial distortion in empirical estimates of subsequent econometric models. Our available illustrative sample of data is far from ideal for truly detailed utility-theoretic demand modelling of environmental values. Nevertheless, the persistent appearance of nontrivial biases in key parameter estimates, even in a selection of simplistic demand models, certainly leads one to suspect that analogous biases would be possible in more sophisticated demand and/or utility models. This inference can readily be extended beyond the boundaries of environmental valuation to all types of other studies using mail survey data.

The main contribution of this paper is its demonstration that reliance on little more than the zip code information available for each household in the target sample allows one to reconstruct a selection of variables that can potentially be used in a response/nonresponse discrete choice model. Since our surveys ask respondents to consider environmental goods at specific identifiable geographical locations, distance is likely to be related to the salience of the good. We also rely on the 1990 Census, aggregated to the level of zip codes, to provide crude measures of the sociodemographic characteristic of each potential respondent's neighborhood. We can also control for membership in different types of subsamples. In our example, many of these Census variables are shown to make a statistically significant contribution to explaining a potential respondent's propensity to complete and return our questionnaire. Similar geographic or sociodemographic considerations or convenience samples will be present in many other types of surveys and the implications of our example extend to any study using these data.

It would have been advantageous to have access to additional variables for this study. In some cases of environmental valuation, for example, it may be possible to solicit from each state information on the numbers of fishing licenses per zip code and/or the number of licensed boat trailers per zip code (for example).[13] One must assume that the direct mail advertising industry also knows a lot about the preferences of US residents by zip code. While such data are unlikely to be free, it may be possible to acquire data on the number of subscribers to certain publications by zip code, or membership in certain organizations. A selection of such zip code frequency variables could paint an even more informative picture of probable survey topic salience.

In order to focus attention on the problem of nonresponse bias in survey research concerning the demand for environmental goods, our first featured examples employ a very simplified model of demand for a aggregates of similar environmental goods. Clear nonresponse biases can show up here. Next, we resort to demand models for individual localized goods. This second class of models is most theoretically satisfying, and provides evidence that the basic implications of demand theory are met. However, the difficulties we experienced in getting these models to converge leads us to offer them as supplementary evidence, rather than to present them as our main results.

Implementing a model of response/nonresponse requires only that sufficient geographic information be retained for the entire intended sample. Researchers must also have access to recent Census data at a corresponding level of aggregation, as well as relevant distance-calculating software.

What is our recommendation? Any researcher using mail survey data should be strongly encouraged to *plan for*, and then to undertake, explicit modelling of response/nonresponse to his or her survey instrument in a manner analogous to that presented here. This is especially important if one expects considerable heterogeneity in the sociodemographic characteristics of potential respondents, or if geographical proximity to the place(s) or object(s) featured in the subject matter of the survey varies substantially across potential respondents. It is also important if there are different versions of the survey, or if portions of the working sample consist of non-random convenience samples appended to a base sample that is reasonably representative. The key insight is that *without* formal nonresponse modelling and correction, the default presumption must be that substantial nonresponse biases could easily be present in any statistical work conducted using only a sample of mail survey *respondents*. These biases can distort not only estimates of the level of demand in the population, but also estimates of the degree of substitutability among goods and overall welfare calculations.

REFERENCES

Ayal, I. and J. Hornik (1986) "Foreign Source Effects and Response Behavior in Cross-National Mail Surveys," *International Journal of Research in Marketing*, 3, pp. 157-67.

Callaway, John M., Shannon Ragland, Sally Keefe, Trudy Ann Cameron, W. Douglass Shaw (1995) *Columbia River System Operation Review Recreation Impacts: Demand Model and Simulation Results*, report prepared for U.S. Army Corps of Engineers, Portland Oregon, by Hagler Bailly Consulting, Inc., Boulder, CO.

Cameron, T.A., W.D. Shaw, S.R. Ragland, J.M. Callaway, and S. Keefe (1996) "Using Actual and Contingent Behavior Data with Differing Levels of Time Aggregation to Model Recreation Demand," *Journal of Agricultural and Resource Economics* (forthcoming).

Dillman, D.A. (1978) *Mail and Telephone Surveys: The Total Design Method*, New York: Wiley.

Dillman, D.A. J.J. Dillman, and C.J. Makela (1984) "The Importance of Adhering to Details of the Total Design Method (TDM) for Mail Surveys," in D.C. Lockhart (ed.) *Making Effective Use of Mailed Questionnaires*. New Directions for Program Evaluation, no. 21. San Francisco: Jossey-Bass.

Edwards, S.F. and G.D. Anderson (1987) "Overlooked Biases in Contingent Valuation Surveys: Some Considerations," *Land Economics*, 63(2), pp. 168-178.

Englin, J., J.S. Shonkwiler, and W.D. Shaw (1996, in progress) "Count Models with Self-Selectivity Corrections: An Application to Recreational Demand Modelling," manuscript, Department of Applied Economics and Statistics, University of Nevada, Reno, NV 89557.

Fisher, Mark R. (1996) "Estimating the Effect of Nonresponse Bias on Angler Surveys," *Transactions of the American Fisheries Society*, 125, 118-126.

Fox, R.J., M.R. Crask, and J. Kim, "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response," *Public Opinion Quarterly*, 52, pp. 467-491.

Goyder, J. (1985) "Nonresponse on Surveys: A Canada-United States Comparison," *Canadian Journal of Sociology*, 10(3), pp. 231-251.

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

Greene, W.H. (1995) *LIMDEP: User's Manual and Reference Guide, Version 7.0.* Bellport, NY: Econometric Software, Inc.

Heberlein, T., and R. Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature," *American Sociological Review*, 43, pp. 447-462.

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

Hellerstein, D., D. Woo, D. McCollum, and D. Donnelly (1993) *ZIPFIP: A ZIP and FIPS Database Users Manual*, Economic Research Service (ERA-NASS), 340 Victory Drive, Herndon, VA 22070 (1-800-999-6779).

Jaffe, E.D. (1982) "The Efficacy of Mail Surveys in Developing Countries--The Case of Israel," *European Research*, 10, pp. 102-104.

Jobber, D., H. Mizra, and K.H. Wee (1991) "Incentives and Response Rates to Cross-National Business Surveys: A Logit Model Analysis," *Journal of International Business Studies*, pp. 711-721.

Jobber, D., and J. Saunders (1988) "An Experimental Investigation into Cross-National Mail Survey Response Rates," *Journal of International Business Studies*, pp. 483-489.

McDaniel, S.W., C.S. Madden, and P. Verille, (1987) "Do Topic Differences Affect Survey Non-Response?" *Journal of the Market Research Society*, 29 (1), pp. 55-66.

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

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

Whitehead, J.C. (1991) "Environmental Interest Group Behavior and Self-Selection Bias in Contingent Valuation Mail Surveys," *Growth and Change*, 22(1), pp. 10-21.

Whitehead, J.C., P.A. Groothuis, and G.C. Blomquist (1993) "Testing for Non-Response and Sample Selection Bias in Contingent Valuation; Analysis of a Combination Phone/Mail Survey," *Economics Letters* **41**, pp. 215-220.

Wiseman, F. and M. Billington (1984) "Comment on a Standard Definition of Response Rates," *Journal of Marketing Research*, 21, pp. 336-338.

TABLE 1 - Descriptive Statistics[a], Response/Nonresponse Sample (n = 7034)
(means; standard deviations for non-binary variables in parentheses)

VERSION	1 (n = 1428)	2 (n = 1433)	3 (n = 2095)	4 (n = 2078)
OVERALL RESP.	0.3573	= 1 if response sufficiently complete (0 otherwise)		
VERSION RESP.	0.3480	0.3517	0.3518	0.3730
HAV-DIST	0.9414	= 1 if distance data available for all specified waters		
HAV-OTH	0.9663	= 1 if distance data available, 5 nearest "other" waters		
HAV-CENSUS	0.9555	= 1 if 1990 Census data available for zip code		
POP-2	0.4362	= 1 if adjacent county sample (0 otherwise)		
POP-3	0.05530	= 1 if "Phase 1" sample (0 otherwise)		
POP-4	0.08189	= 1 if "known user" sample (0 otherwise)		
POP-5	0.02132	= 1 if Canadian sample (0 otherwise)		
VERSION	-	0.2037	0.2978	0.2954
DIST-1	0.2386 (1.061)	0.3616 (0.9550)	0.8134 (1.481)	0.5058 (0.9349)
DIST-2	0.3467 (1.200)	0.3334 (0.9204)	0.4664 (0.8796	0.7981 (1.437))
DIST-3	0.2702 (1.088)	0.2892 (0.8043)	0.7835 (1.393)	0.6676 (1.233)
DIST-4	0.2579 (1.062)	0.4300 (1.210)	-	0.7302 (1.344)
DIST-OTHER	0.2090 (1.048)	0.1979 (0.6640)	0.3039 (0.6683)	0.3881 (0.7705)
PLANGIS	0.004839 (0.006962)	= Proportion of zip code population language-isolated		
PPUBINC	0.02456 (0.01493)	= Proportion of zip code population on public assistance		
PURBAN	0.5041 (0.3978)	= Proportion of zip code population in urban area		
PAGOCC	0.02690 (0.03265)	= Proportion of zip code population in ag., fishing, or forestry-related occupations		
PSSINC	0.1023 (0.04508)	= Proportion of zip code population on social security		

[a] Identity of the specific waters to which distances are measured differs across survey versions:
Version 1: DIST-1 = Hungry Horse Reservoir, DIST-2 = Lake Pend Oreille, DIST-3 = Lake Koocanusa, DIST-4 = Kootenai River, DIST-OTHER = other waters in the Version 1 region;
Version 2: DIST-1 = Dworshak Lake, DIST-2 = Clearwater River, DIST-3 = Lower Granite Lake, DIST-4 = Lake Pend Oreille, DIST-OTHER = other waters in the Version 2 region;
Version 3: DIST-1 = Lake Roosevelt, DIST-2 = Lake Umatilla, DIST-3 = Lower Granite Lake, DIST-OTHER = other waters in the Version 3 region;
Version 4: DIST-1 = Lake Roosevelt, DIST-2 = Dworshak Lake, DIST-3 = Lower Granite Lake, DIST-4 = Lake Pend Oreille, DIST-OTHER = other waters in the Version 4 region.

TABLE 2 - Stage A: Probit Model for Survey Response/Nonresponse (n = 7034)
(point estimates; asymptotic t-ratios in parentheses; ** = 5%, * = 10% level)

VARIABLE	VERSION 1[a]	VERSION 2	VERSION 3	VERSION 4
CONSTANT	-5.4090 (-0.101)	-	-	-
HAV-DIST	-0.94358 (-8.038)**	-	-	-
HAV-OTH	5.8271 (0.109)	-	-	-
HAV-CENSUS	-0.041423 (-0.255)	-	-	-
POP-2	0.051603 (0.935)	-	-	-
POP-3	0.19816 (2.766)**	-	-	-
POP-4	0.45169 (5.787)**	-	-	-
POP-5	0.056020 (0.001)	-	-	-
VERSION DUMMY	-	-0.011991 (-0.133)	0.33752 (2.073)**	0.10617 (0.956)
DIST-1	0.029962 (0.251)	-0.012718 (-0.061)	-0.10331 (-1.242)	-0.13793 (-1.620)
DIST-2	0.00064882 (0.014)	-1.4318 (-2.676)**	-0.14959 (-1.464)	0.33916 (1.353)
DIST-3	0.24446 (0.898)	1.5193 (3.020)**	-0.017056 (-0.302)	-0.31424 (-1.164)
DIST-4	-0.22852 (-1.190)	0.15348 (3.289)**	-	0.16944 (1.325)
DIST-OTHER	-0.089274 (-2.334)**	-0.11004 (-1.284)	0.30916 (1.727)*	-0.27913 (-1.515)
PLANGIS	-5.9098 (-2.056)**	-	-	-
PPUBINC	-4.2524 (-3.395)**	-	-	-
PURBAN	0.14932 (2.449)**	-	-	-
PAGOCC	0.73584 (1.042)	-	-	-
PSSINC	1.0420 (2.392)**	-	-	-
Log $\mathcal{L}$	-4341.5			

[a] Waters corresponding to each "numbered distance" differ across versions.

TABLE 3 - Descriptive Statistics, Respondents Only
(means; standard deviations for non-binary variables in parentheses)

VARIABLE	VERSION 1 (n = 497)	VERSION 3 (n = 737)	VERSION 4 (n = 775)	DESCRIPTION
TOTAL-TRIPS	10.12 (17.43)	3.550 (7.532)	5.737 (12.80)	number of water-based recreation trips in 1993
SINGLE-SITE TRIPS	0.2676 (1.330)	-	0.2787 (1.485)	Version 1; trips to Lake Koocanusa Version 4; trips to Lake Pend Oreille
HAV-DIS	0.8350 (0.3715)	-	0.9458 (0.2265)	distance data available to all relevant waters
AVG-DIST	0.07074 (0.1242)	0.1037 (0.1105)	0.1199 (0.08057)	Average distance to nearest five waters of any description
HAV-INC	0.8330 (0.3734)	0.8521 (0.3552)	0.8490 (0.3582)	= 1 if respondent provided income data (0 otherwise)
INC	0.02945 (0.02885)	0.03193 (0.02860)	0.03331 (0.02969)	Respondent average monthly income in $ 100,000
HAV-AGE	0.9678 (0.1767)	0.9647 (0.1846)	0.9523 (0.2134)	= 1 if respondent provided age data (0 otherwise)
AGE	5.161 (1.734)	5.089 (1.821)	4.937 (1.878)	Respondent age (in tens of years)
FISH-LICENSE	0.6016 (0.4901)	0.4301 (0.4954)	0.4052 (0.4912)	= 1 if respondent holds a current fishing license (0 otherwise)
OWN-BOAT	0.5171 (0.5002)	0.3636 (0.4814)	0.3858 (0.4871)	= 1 if respondent owns a boat (0 otherwise)
IMR	0.9747 (0.2985)	1.025 (0.1218)	0.9672 (0.1812)	= fitted inverse Mill's ratio from first-stage probit model

TABLE 4 - Stage B1: TOBIT ESTIMATES: With and Without Nonresponse Selectivity Correction
(point estimates; Murphy-Topel corrected asymptotic standard errors in parentheses)
Dependent Variable: Total Number of Trips to *Any* Water in Region

	VERSION 1 (n = 497)		VERSION 3 (n = 737)	
VARIABLE	Corrected	Uncorrected	Corrected	Uncorrected
CONSTANT	9.2763 (1.284)	-0.93730 (-0.144)	12.695 (2.456)**	-7.0195 (-2.232)**
HAV-DIS	0.53492 (0.091)	-16.707 (-6.776)**	-	-
AVG-DIST	-38.778 (-4.116)**	-43.334 (-4.751)**	-26.359 (-5.735)**	-20.916 (-4.647)**
HAV-INC	0.75177 (0.260)	0.66804 (0.230)	1.5582 (0.955)	1.1215 (0.686)
INC	11.290 (0.321)	-1.0807 (-0.031)	6.2325 (0.325)	6.8328 (0.354)
HAV-AGE	32.242 (4.463)**	33.566 (4.663)**	8.2059 (2.386)**	9.0126 (2.611)**
AGE	-3.7163 (-5.700)**	-3.7889 (-5.779)**	-1.3805 (-4.198)**	-1.4887 (-4.494)**
FISH-LICENSE	7.6388 (3.657)**	8.0040 (3.816)**	3.8794 (3.602)**	4.0515 (3.730)**
OWN-BOAT	7.1948 (3.558)**	7.6478 (3.768)**	6.7424 (6.191)**	7.3264 (6.696)**
IMR	-24.809 (-3.235)**	-	-18.579 (-4.673)**	-
SIGMA	18.840 (25.559)**	18.995 (25.513)**	11.240 (24.704)**	11.388 (24.631)**
Log $\mathcal{L}$	-1566.8	-1572.1	-1532.7	-1543.7
Fitted trips	23.91	12.47	13.05	6.275

TABLE 5 - Stage B2: TOBIT ESTIMATES: With and Without Nonresponse Selectivity Correction
(point estimates; Murphy-Topel corrected asymptotic standard errors in parentheses)
Dependent Variable = Number of Trips to **Specific Single Water** in Region

VERSION	1 (n = 497) Lake Koocanusa (W3)		4 (n = 775) Lake Pend Oreille (W4)	
VARIABLE[a]	Corrected	Uncorrected	Corrected	Uncorrected
CONSTANT	7.7610	-0.36457	-2.0284	-20.373
	(1.541)	(-0.086)	(-0.007)	(-0.064)
HAV-COST	-9.3358	-8.7272	6.7455	11.155
	(-2.471)**	(-2.258)**	(0.022)	(0.035)
RTC-W1	28.348	35.548	29.685	43.012
	(2.855)**	(3.541)**	(3.191)**	(4.716)**
RTC-W2	23.307	26.335	34.774	40.528
	(1.524)	(1.666)*	(0.938)	(1.234)
RTC-W3	**-58.351**	**-61.149**	-23.500	-27.670
	(-2.858)**	**(-2.965)****	(-0.640)	(-0.856)
RTC-W4	0.19742	-5.4258	**-40.798**	**-57.414**
	(0.008)	(-0.212)	**(-4.781)****	**(-6.418)****
DIST-OTHER	-0.50069	-0.51961	-0.64864	-0.99794
	(-1.509)	(-1.526)	(-0.275)	(-0.455)
HAV-INC	-0.20109	0.21138	-1.5609	-0.54145
	(-0.116)	(0.121)	(-0.654)	(-0.235)
INC	24.844	18.336	-56.832	-65.482
	(1.094)	(0.792)	(-1.287)	(-1.535)
HAV-AGE	3.6719	3.6488	3.7734	3.4295
	(1.104)	(1.075)	(1.020)	(0.963)
AGE	-0.73097	-0.81161	-0.62715	-0.70838
	(-1.989)**	(-2.118)**	(-1.292)	(-1.504)
FISH-LICENSE	-0.34727	0.20197	1.0845	0.68323
	(-0.306)	(0.177)	(0.794)	(0.502)
OWN-BOAT	3.1394	3.5133	1.8448	2.9300
	(2.587)**	(2.813)**	(1.304)	(2.061)**
IMR	-7.0387	-	-16.170	-
	(-2.804)**		(-3.906)**	
SIGMA	5.4358	5.6438	6.2956	6.5700
	(8.275)**	(8.220)**	(8.039)**	(7.962)**
Max Log $\mathcal{L}$	-216.66	-220.75	-186.91	-196.64

[a]Round-trip costs ($'00) for different versions: Version 1: RTC-W1 = Hungry Horse Reservoir, RTC-W2 = Lake Pend Oreille, RTC-W3 = Lake Koocanusa, RTC-W4 = Kootenai River; Version 4: RTC-W1 = Lake Roosevelt, RTC-W2 = Dworshak Lake, RTC-W3 = Lower Granite Lake, RTC-W4 = Lake Pend Oreille.

TABLE A-1

Candidate Census Variables for Response/Nonresponse Probit Submodel

ACRONYM	CONSTRUCTION FROM STANDARD CENSUS STF3 VARIABLES	INTERPRETATION
PERSONS	P1_1	Population of zip code area
PURBAN	P6_1/P1_1	Proportion urban
PWHITE	P8_1/P1_1	Proportion White
PBLACK	P8_2/P1_1	Proportion Black
PAMIN	P8_3/P1_1	Proportion Native American
PASIAN	P8_4/P1_1	Proportion Asian
POTHER	P8_5/P1_1	Proportion other ethnicity
PLANGIS	(P29_2+P29_4+P29_6)/P1_1	Proportion language-isolated
PLTERM	P43_1/P1_1	Proportion long-term resident (same dwelling in 1985)
PCOLL	(P60_6+P60_7)/P1_1	Proportion college-educated
PAGIND	P77_1/P1_1	Proportion in agriculture, fishing or forestry industries
PAGOCC	P78_9/P1_1	Proportion in agriculture, fishing or forestry occupations
PSSINC	P94_1/P1_1	Proportion on social security income
PPUBINC	P95_1/P1_1	Proportion on public assistance income
PRETINC	P96_1/P1_1	Proportion with retirement income
INCM	P80A_1/1000	Median household income ($'000)
RENT	H43A_1/1000	Median rental rate ($'000)
VALUE	H61A_1/1000	Median house value ($'000)

Appendix 1 - **Tobit Model with Sample Selection**

A little intuition will help with the development of the appropriate log-likelihood function. The domain of the joint density function can be partitioned into three distinct regions. The first region is characterized by $y_i^* > 0$ and $q_i > 0$ (respondents with nonzero observed trips). The second region has $y_i^* > 0$ and $q_i = 0$ (respondents with zero trips). The third region has $y_i^* < 0$ and thus q_i unknown (the nonrespondents).

For observations in the first region, the joint density can be conveniently expressed as the marginal density of q_i (observed) times the conditional density of y_i^* given the value of q_i. The random variable q_i is $N(z_i'\gamma, \sigma^2)$ and $f(y_i^* | q_i)$ is also normal with mean $x_i'\beta + \rho[(q_i - z_i'\gamma)/\sigma]$ and variance $(1 - \rho^2)$, since the variance of y_i^* is normalized to unity. The term for the marginal distribution of q_i will look like the ordinary maximum likelihood regression formula. The term for the conditional distribution will look like the term for a conventional MLE probit model for the positive domain of the latent variable. For this region, then, the contribution of one observation to the log-likelihood function is:

$$\log \mathcal{L}_{1i} = \{ -.5 \log(2\pi) - \log \sigma - .5 [(q_i - z_i'\gamma)^2/\sigma^2] \} + \log[1 - \Phi(R_i)]$$

where $R_i = - \{ x_i'\beta + \rho[(q_i - z_i'\gamma)/\sigma] \} / (1 - \rho^2)^{.5}$.

For the second region, we assume that all values of $q_i^* < 0$ are manifested in the observed data as $q_i = 0$. Here, we must use the appropriate cumulative density associated with the bivariate normal distribution. If Φ_2 (a,b,ρ) denotes the cumulative standard bivariate normal density function evaluated up to limits a and b, the log-likelihood terms for observations in this second region are given by:

$$\log \mathcal{L}_{2i} = \log[\Phi_2 (x_i'\beta, -z_i'\gamma/\sigma, -\rho)$$

For the third region of the domain of the joint density, all that is known is that $y_i^* < 0$, so we use the simple marginal distribution of y_i^*, employing a term like the one that applies to the negative domain of a conventional probit log-likelihood:

$$\log \mathcal{L}_{3i} = \log \Phi [-x_i'\beta].$$

Putting all three of these terms together, the full log-likelihood objective function can be expressed as:

$$\max_{\beta,\gamma,\sigma,\rho} \quad \log \mathcal{L} = \Sigma_{y_i=1, q_i>0} \log \mathcal{L}_{1i} + \Sigma_{y_i=1, q_i=0} \log \mathcal{L}_{2i} + \Sigma_{y_i=0} \log \mathcal{L}_{3i}$$

Appendix 2 - **Murphy-Topel Corrected Second-Stage Variance-Covariance Matrix**

Since FIML estimates and the desirable variance-covariance matrix cannot be attained in this application, we adopt the framework of Murphy and Topel (1985) in order to correct the second-stage variance-covariance matrix in our two-stage Tobit model. Let $N = nr + nn$ be the total number of observations in the target sample, with nr being the number of respondents and nn the number of nonrespondents. Let $I_i = 1$ if $q_i > 0$, and $I_i = 0$ if $q_i = 0$. The two separate log-likelihood functions employed in the two-stage method are:

$$\log\mathcal{L}_1 = \Sigma_N\, y_i \log \Phi(x_i'\beta) + (1 - y_i) \log [1- \Phi(x_i'\beta)], \text{ and}$$

$$\log\mathcal{L}_2 = \Sigma_{nr}\, (-I_i/2)[\log(2\pi) + \log\sigma^2 + ((q_i - z_i'\gamma - \gamma_\lambda\lambda_i)/\sigma)^2] +$$
$$(1 - I_i) \log[1 - \Phi((z_i'\gamma + \gamma_\lambda\lambda_i)/\sigma)],$$

where $\lambda = \phi(x_i'\beta)/\Phi(x_i'\beta)$. If we now define $\theta = (\gamma', \gamma_\lambda, \sigma)'$, Murphy and Topel (1985) demonstrate that the correction formula for the second stage standard error estimates involves four matrices:

$$R_1 = -E[(\partial\log\mathcal{L}_1/\partial\beta)(\partial\log\mathcal{L}_1/\partial\beta'] = -E[\partial^2\log\mathcal{L}_1/\partial\beta\partial\beta']$$

$$R_2 = -E[(\partial\log\mathcal{L}_2/\partial\theta)(\partial\log\mathcal{L}_2/\partial\theta'] = -E[\partial^2\log\mathcal{L}_2/\partial\theta\partial\theta']$$

$$R_3 = -E[(\partial\log\mathcal{L}_2/\partial\beta)(\partial\log\mathcal{L}_2/\partial\theta'] = -E[\partial^2\log\mathcal{L}_2/\partial\beta\partial\theta']$$

$$R_4 = -E[(\partial\log\mathcal{L}_1/\partial\beta)(\partial\log\mathcal{L}_2/\partial\theta']$$

The matrices R_1 and R_2 can be replaced by the inverses of the uncorrected estimators for the asymptotic covariance matrices for the first stage probit and the second stage Tobit coefficients, respectively. Matrices R_3 and R_4 must be specially constructed. Based on the two-stage log-likelihood expressions, we have the vectors of derivatives:

$$\partial\log\mathcal{L}_1/\partial\beta = \Sigma_N \{ y_i\lambda_i - (1-y_i)\, \phi(x_i'\beta)/(1 - \Phi(x_i'\beta)) \}\, x_i$$

$$\partial\log\mathcal{L}_2/\partial\gamma = \Sigma_{nr}\, (1/\sigma) [I_i\, ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))]\, z_i$$

$$\partial\log\mathcal{L}_2/\partial\gamma_\lambda = \Sigma_{nr}\, (1/\sigma) [I_i\, ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))]\, \lambda_i$$

$$\partial\log\mathcal{L}_2/\partial\beta = \Sigma_{nr} \{ [I_i\, ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))] *$$
$$(\gamma_\lambda/\sigma)\, [(x_i'\beta)\, \lambda_i - \lambda_i^2] \}\, x_i.$$

and the scalar derivative:

$$\partial\log\mathcal{L}_2/\partial\sigma = \Sigma_{nr}\, (1/\sigma) \{I_i [((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma)^2 - 1] +$$
$$(1-I_i)\, ((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)\;\; \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))$$

The complete vector $\partial\log\mathcal{L}_2/\partial\theta$ is constructed from $((\partial\log\mathcal{L}_2/\partial\gamma)',(\partial\log\mathcal{L}_2/\partial\gamma_\lambda)',\partial\log\mathcal{L}_2/\partial\sigma)')$. The outer products of the appropriate vectors of derivatives are used to calculate R_3 and R_4.

Once the component R matrices have been calculated, the corrected variance-covariance matrix for the parameter vector of the second-stage tobit model will be $\sqrt{n}(\theta\text{-}\theta) \sim N(0,\Sigma)$, where

$$\Sigma = R_2^{-1} + R_2^{-1}[R_3'R_1^{-1}R_3 - R_4'R_1^{-1}R_3 - R_3'R_1^{1}R_4]\, R_2^{-1}.$$

Appendix 3 - **Consistent Estimation of Tobit Conditional Error Variance**

As in Greene (1993, p. 711), define:

$$\alpha_y = -x_i'\beta$$

$$\lambda_i = \phi(\alpha_y)/\Phi(\alpha_y)$$

$$\delta_i = \lambda_i [\lambda_i + x_i'\beta]$$

For each observation, i, the true error variance would be $\sigma_i^2 = \sigma_\epsilon^2(1 - \rho^2\delta_i)$. The average variance for the sample errors would converge in the limit to:

$$\text{plim } (1/n)\ \sigma_i^2 = \sigma_\epsilon^2(1 - \rho^2\delta^*)$$

where δ^* is the mean of the δ_i values. The maximum likelihood second stage Tobit algorithm provides an estimate, σ_0^2, of the desired quantity plim $(1/n)\sigma_i^2$.

Another necessary component is provided by the square of the coefficient on the inverse Mill's ratio term, λ_i. Let this coefficient be denoted γ_λ. We can use the result that plim $\gamma_\lambda^2 = \rho^2\sigma_\epsilon^2$. The first-stage probit model provides individual estimates of δ_i and plim $(1/n)\Sigma_i\delta_i = \delta^*$. Finally, we can generate a consistent estimator for the desired σ_ϵ^2 using the formula:

$$\sigma_\epsilon^2 = \sigma_0^2 + \delta^*\ \gamma_\lambda^2.$$

Tobit models produce estimated parameters, which, when employed in linear combination with the explanatory variables, produce fitted values of the Tobit "index." This index is the conditional expected value of the latent q_i^* variable. Fitted values of $q_i^* < 0$ are interpreted as zero fitted values of the observable trips variable, q_i. To determine the expected number of trips for a given vector of explanatory variables, the index must therefore be manipulated somewhat further. The expected number of trips is given by the probability of positive trips times the expected number of trips, conditional on trips being positive. This conditional probability depends upon the estimate of the error variance. One must be careful to use σ_ϵ^2 rather than the value of σ_0^2 produced automatically by the second-stage ordinary Tobit estimator.

Recall that the expected value of a standard normal random variable truncated below at c is given by $\phi(c)/[1 - \Phi(c)]$. Thus the implied conditional density function for individual i, truncated below at zero, will have a mean of

$$(z_i'\gamma + \gamma_\lambda\lambda_i) + \sigma_\epsilon [\phi((z_i'\gamma + \gamma_\lambda\lambda_i)/\sigma_\epsilon)/\Phi((z_i'\gamma + \gamma_\lambda\lambda_i)/\sigma_\epsilon)].$$

The fitted $E[q_i]$ will be this value multiplied by the fitted probability of positive trips for this individual: $\Phi((z_i'\gamma + \gamma_\lambda\lambda_i)/\sigma_\epsilon)$. To simulate circumstances with no non-response bias, we can set $\gamma_\lambda\lambda_i = 0$ for all observations.

ENDNOTES

1. Some subsequent research concerning the prescriptions in Dillman (1978) is described in Dillman et al. (1984). Twenty-nine different elements of the "total design method" were either adhered-to or not for samples taken from eleven different states in the U.S. and the consequences for response-rates evaluated.

2. For a number of other countries, analogous methods are potentially feasible, depending on the availability of similar types of Census and distance data.

3. Wiseman and Billington (1984) address the issue of standardizing the definition of "response rate" in applied statistics. In the present study, the response rate is defined as the number of usable returned questionnaires divided by the total number of questionnaires mailed out. No adjustments are included for "returned undeliverable" or other exclusions that are occasionally allowed before making this computation.

4. Saltwater recreation opportunities may be viewed by some households as substitutes for the freshwater recreation opportunities they are being asked to consider. Saltwater resources were not considered in this survey.

5. Trips to any one water mentioned in any one survey version are much more sparse than total trips to all waters for that version, so not all submodels converge.

6. In a more-elaborate model, the specification could distinguish between complete non-response and unusable responses. However, this would require a trivariate joint density for FIML estimation.

7. Since distances to waters may be negatively correlated with distances to major urban areas, which have not been controlled for in our models, these results may be open to different interpretations.

8. Varying degrees of multicollinearity among some of the Census variables exist, but the purpose of the first-stage response/nonresponse model is to predict response probabilities, so this problem is not too troubling. The important result is that there is significant systematic variation in response probabilities.

9. Keeping in mind that the aggregate total of all types of water-based recreation is being modelled in this illustrative application (sight-seeing, camping, picnicking, etc., not just fishing trips, for example), the fishing license and boat-ownership dummy variables are less likely to be completely jointly determined with the dependent TRIPS variable.

10. The implied point value of ρ exceeds one in absolute value. In finite samples, and without parameter restrictions, this is possible.

11. We owe this eminently sensible explanation to Michael Hanemann.

12. Englin et al. (1996) consider Poisson-based selectivity models.

13. We use individual data on fishing licenses and boat ownership available for the respondent sample in the demand portion of our two-stage modelling exercise, but analogous zip code level variables could also contribute substantially to capturing the salience of water-recreation issues to the overall target population.

CONTINGENT VALUATION OF INSTREAM FLOWS IN NEW MEXICO: TESTS OF SCOPE, INFORMATION AND TEMPORAL RELIABILITY

Robert P. Berrens*
Carol L. Silva
David Brookshire
Alok K. Bohara
Philip Ganderton
Hank Jenkins-Smith

University of New Mexico

ABSTRACT

This study uses the contingent valuation (CV) method to investigate the nonmarket benefits of protecting minimum instream flows in New Mexico. The original dichotomous choice CV telephone survey was conducted in February 1995, and used a voluntary contribution trust fund format. Using the same 2 x 2 experimental design, CATI (computer-assisted telephone interview) system and sampling frame of New Mexico residents, the survey was then replicated in February 1996. The combined data sets are used to conduct tests of: (i) sensitivity in valuation responses to a change in the scope of the good; (ii) sensitivity in valuation responses to information about the collective nature of providing the good; and (iii) the temporal reliability of results.

Key Words: contingent valuation, instream flows, New Mexico

---

* No lead authorship is assigned. Berrens, Bohara, Brookshire and Ganderton are in the Department of Economics; Jenkins-Smith and Silva are with the Institute for Public Policy, UNM.

Partial support for this research was provided by a National Science Foundation/Environmental Protection Agency (NSF/EPA) grant. All errors and interpretation of the results are solely the responsibility of the authors.

Address correspondence to R. Berrens at 1915 Roma N.E., Albuquerque, NM 87131; email: rberrens@unm.edu.

INTRODUCTION

This empirical study, composed of two separate samples, uses the survey-based contingent valuation (CV) method to estimate nonmarket values for protecting minimum instream flows in New Mexico (NM) rivers. The CV method is a valuable tool for measuring the value of nonmarket environmental goods, and can provide important information in the larger decision making system. Spurred by changes in natural resource damage law, the method has undergone intense scrutiny. Continued refinement requires formal hypothesis testing and the accumulation of evidence. The telephone survey instrument used here includes a dichotomous choice (DC) elicitation format, voluntary contribution trust fund payment vehicle, and split-sample treatments for three hypotheses tests.

HYPOTHESES TESTS

First, we test for split-sample sensitivity in valuation responses to a change in the scope of the good. There are a variety of nesting and sequencing phenomenon loosely referred to as scope, part-whole, and embedding effects (Brown and Duffield, 1995). Following Carson and Mitchell's (1995) categorization, we conduct an external (split-sample) scope test of *component sensitivity* for geographically nested goods. This corresponds to Kahneman and Knetsch's (1992) *perfect embedding*.

Specifically, the *test of scope* compares values for protecting minimum instream flows for a single endangered fish in a 170 mile river stretch (the silvery minnow in the Middle Rio Grande) versus the larger composite good of protecting minimum instream flows on four major NM rivers with 11 threatened and endangered fish species.

Second, we test for split-sample sensitivity in valuation responses to the inclusion of information about the collective nature of the providing the public good. The specific *test of information* follows Green et al. (1994) in including a brief reminder statement immediately prior to the valuation question on the group-size supporting the public good. In combination with the scope test, this test is motivated by a series of recent papers by Kahneman and colleagues. They argue that there are two competing models for how individuals answer valuation questions: the purchase and contribution models (Green et al, 1994.; Kahneman et al., 1993; and Kahneman and Ritov, 1994).

The purchase model that underlies much of the CV literature constructs a hypothetical market wherein the respondent compares two states of the world and provide the income adjustment that makes her indifferent to a posited change in an environmental good. Willingness to pay (or be paid) responses are interpreted as valid measures of welfare change. The purchase model emphasizes the acquisition of a precisely demarcated good.[1]

The contribution model posits that individuals view public goods provision as good causes that need support (Kahneman and Knetsch, 1992; Kahneman et al., 1993). Under the contribution model, WTP expresses an attitude to a public good or general cause, and as such entails generally low sensitivity to changes in scope. Valuation responses that express general attitudes may be the source of intrinsic satisfaction and "warm glow" effects. Further, detailed information about the good, or alternative public goods, may not be important and sketchy descriptions or "headlines" may be sufficient to elicit the same expression of attitudes (Kahneman and Ritov, 1994). However, information about the number of potential contributors may matter in that it affects perceived social norms on acceptable levels of contributions (Green et al., 1994), whereas the purchase model would be invariant to such a reminder.

In order to advance CV research the purchase and contribution models must produce competing hypotheses. Carson and Mitchell (1995) review and provide evidence that a well-defined CV survey instrument will commonly show sensitivity to changes in the scope of the good, including those with expected nonuse values. Smith (1996) provides empirical evidence that DC-CV can discriminate

---

[1] A measure of support for the purchase model is the ubiquitous finding of significant negative bid responsiveness in DC models, as is other empirical evidence of construct validity (e.g., Smith, 1996).

between different objects of choice with public good characteristics, and argues that this is evidence against the contribution model and "headline" method of Kahneman and Ritov (1994).

In the initial empirical test of the group-size reminder, Green et al. (1994) identify highly significant reminder effects that *lowered* valuations for several public goods by 50 percent or more. This negative response effect, using an open-ended elicitation format and in-person interviews, was found when describing both one million and 10 million potentially contributing households (Green et al., 1994); it was also stable across payment vehicles (taxes and voluntary contributions). Using a dichotomous choice format with a telephone survey instrument, the group-size reminder is investigated here *jointly* with the test of scope.

Third, as a *test of temporal reliability* of results, we replicate the original telephone survey, with the same tests of scope and information, one year later. This follows the specific suggestion of Arrow et al. (1993) to reduce "time dependent measurement noise" by averaging across samples drawn at different points in time from the same population (and see Carson et al., 1995). It also should be differentiated from test-retest type approaches where the same set of respondents is resurveyed. The original 1995 telephone survey and DC-CV modeling results are described in Berrens et al. (1996).

BACKGROUND INFORMATION

Instream Flow Protection in New Mexico

Instream flow is the flow of water in its natural channels without diversion. Maintenance of instream flow is desirable to protect and enhance recreation, water quality, and biodiversity. In fully appropriated river systems, instream flow protection may conflict with diversionary uses of water. As elsewhere in the West, the struggle over water allocation has led to careful scrutiny of New Mexico water law, including concern for the protection of instream flows (Bokum et al., 1992; Dumars and Minis, 1989). While transferable, prior appropriation rights are usufructuary, and water must be put to beneficial use, or the right is subject to revocation. New Mexico does not recognize instream flows as a beneficial use of water, and has been politically resistant to any change in the status quo (Bokum et al., 1992; DeYoung, 1993).

Given that beneficial use requires that water be diverted from the streambed, voluntary private market transfers to provide instream flows are unavailable in NM, and generally restricted in most western states. Elsewhere in the West a variety of alternative protection actions have been explored (Bokum et al., 1992, McNalley and Matthews, 1995). In some states a single public agency may purchase water rights to protect instream flows, typically restricted to some minimum requirement.

The failure to protect instream flows is a significant cause of accumulating ecological evidence of degraded riparian ecosystems in NM (Bestgen and Platania, 1991; Crawford et al., 1993, Rinne and Platania, 1995). As a prominent example, in August 1994, the silvery minnow (*Hybnognathus amarus)* was listed as an endangered species by the U.S. Fish and Wildlife Service (USFWS).[2] This tiny fish (approximately 3 1/2") is now found in five percent of its original habitat – relegated to the 170 miles composing the Middle Rio Grande.[3] Low flow events are a critical threat to the silvery minnow, which is considered a bio-indicator of the health of warmwater riverine ecosystem in the Middle Rio Grande (Bestgen and Platania, 1991; Crawford et al., 1993). The silvery minnow is one of eleven threatened or endangered freshwater fish species in NM identified by the USFWS in 1994.

---

[2]The draft economic report on the designation of the silvery minnow's critical habitat (mainstem of the Middle Rio Grande) was released in early spring 1996 (USFWS, 1996). All survey data used in this CV study was collected in the interval between the initial listing decision and the release of the draft report, during which no recovery actions were taken. Severe drought and low flow conditions in spring and summer 1996 have exacerbated concerns for saving the fish, and served as the impetus for initial emergency measures aimed at providing minimum flow. However, exact minimum flow targets have yet to be articulated.

[3]The Middle Rio Grande runs from Cochiti Dam south through the greater Albuquerque area and on to Elephant Butte Dam. The region is essentially a fully appropriated system with mean annual flows, at various gauges, in the 1300-1600 cubic feet per second (cfs) range. Annual discharge is highest during the spring runoff between March and June, and lowest from July to November when irrigation demand peaks. In low flow years lower portions of the mainstem of the Middle Rio Grande will actually run dry for extensive periods.

Nonmarket Values for Instream Flows

Empirical evidence on the nonmarket benefits of instream flows comes in a variety of forms and has been thoroughly reviewed elsewhere (Colby, 1990, 1993; Loomis, 1987). Published studies on recreational use values for instream flows continue to accumulate (e.g., Duffield et al., 1992; Harpman et al., 1993; Loomis and Creel, 1992; Ward). However, both Loomis (1987) and Colby (1990) recognize the importance of both use and nonuse values for instream flows. Nonuse values may be especially important for unique environments or endangered species (Colby, 1993).[4]

There are a variety of CV studies on instream flow protection with expected nonuse values. Berrens et al. (1996) found mean WTP values of $29 for minimum instream flow protection for the silvery minnow in the Middle Rio Grande, and $89 for minimum instream flow protection on all major NM rivers. Other CV studies relating directly or indirectly to valuing the protection of instream flows include a variety of both higher and lower values (Brown and Duffield, 1995; Cummings et al. 1994; Sanders et al., 1990).

The use of CV surveys to estimate environmental values will always be subject to concern over the *ex hypothetico* nature of the expressed preferences. An emergent thread in valuation and experimental laboratory research is to seek out ways to somehow calibrate the expressed values with what might be observed in an actual transaction. Arrow et al (1993) recommend dividing CV estimates by two in natural resource damage and liability assessments as a sort of default rule, unless a preferred calibration alternative is identified. In the only test of real versus hypothetical contributions to an instream flow trust fund, Duffield and Patterson (1992) find evidence that hypothetical contributions

[4]It is expected that the nonmarket values for this study will be significantly composed of nonuse values. While a nontrivial portion of the sample may recreate (hiking, birding, etc.) in riparian areas, we make no attempt to decompose total economic value estimates, which may in part reflect current or expected future use. Our focus is on the protection of minimum instream flows (not recreational optimal flows), and endangered and threatened fish species that are not legally targeted by anglers.

may overestimate relative to actual contributions (ranging from 33% to 100% for different groups). The mail survey sample was drawn from licensed anglers using Montana rivers, and characterized by a low response rate. Further, Loomis (1996) notes that such comparisons beg the question of what the appropriate reference point is, given that actual contributions to public goods may subject to systematic undervaluation due to free-riding.

The question of calibration remains an unresolved CV research issue. Arguing that there may never be a single "crucial test" of the method, Randall (1993) notes empirical research must continue to focus on the accumulation of evidence. This includes the mapping of performance characteristics for selected survey instruments, survey modes and experimental designs (Randall, 1993). While part of a larger research program, the present investigation is targeted to the testing of the three specific hypotheses concerning scope, information, and temporal reliability.

SURVEY INSTRUMENT, EXPERIMENTAL DESIGN AND DATA COLLECTION

The original CV survey was part of a regular Quarterly Profile telephone survey of New Mexico, administered in February 1995 (FEB95) by the Survey Research Center of the Institute for Public Policy (SRC-IPP) at the University of New Mexico. The CV section was then replicated in the February 1996 (FEB96) Quarterly Profile.

The Quarterly Profile is an omnibus survey conducted quarterly since 1988. The statewide survey uses a stratified random sampling approach. Proportionate sampling is used within the working ranges of all telephone number prefixes in NM to obtain a minimum target level of 500 completed interviews. For both the February 1995 and 1996 surveys there was an oversample of approximately 170 completed interviews for Bernalillo County (the greater Albuquerque area). Completed surveys averaged 28 minutes in length.

Table 1 provides complete description of response rates and disposition of all numbers called from both the February 1995 and February 1996 Quarterly Profiles. For the February 1995 survey the contact rate was 75% (completed + appointments not completed + refusals + language barriers / total numbers dialed); the cooperation rate was 64% (completed / completed + appointments not completed + refusals); and the refusal rate was 30% (refusal / completed + refusals). For the February 1996 survey the contact rate was 78%; the cooperation rate was 65%; and the refusal rate was 29%.

The survey instrument included attitudinal and perception questions on topics about New Mexico institutions and politics, as well as numerous socio-economic questions. The valuation section of the survey was pre-tested and refined through several iterations; the final version used in both quarterly profiles is presented in Appendix A. Finally, all interviewers completed in multiple trial runs of the entire telephone survey instrument (see Table 1).

The instream flow section of the survey begins by asking some general awareness questions on New Mexico water issues. The text defines beneficial use and instream flows, and identifies some of the benefits (e.g., fish and wildlife, recreation, water quality) and costs (e.g., higher prices, restricted development) of protecting instream flows. Respondents are then told of the number (11) of endangered and threatened fish species in NM, and the four separate rivers (Gila, Pecos, Rio Grande and San Juan) where they are found. The text of a split-sample treatment includes a brief statement identifying the silvery minnow of the Middle Rio Grande as one of the 11 fish species. All respondents are told that protecting endangered fish and their habitat may require protecting minimum instream flow, and that trust funds are used in some states to buy or lease water for such purposes. Prior to the valuation section, respondents are told that they will be asked about the dollar value their household places on protecting instream flows and that there are no right or wrong answers, and then reminded of household budget constraints and available substitutes.

Using the CATI (computer-assisted telephone interview) system, the valuation section employs a 2 x 2 experimental design for split sample hypothesis testing. Ignoring the replication aspect, the two specific hypotheses to be tested are: (1) sensitivity to a change in the scope of the good, and (2) sensitivity to a reminder on the group size (500,000 households) potentially contributing to the provision of the public good. For modeling, scope is hereafter indicated by the dummy variable, SM, where SM=1 indicates the treatment sample that receives the silvery minnow valuation question, and SM=0 indicates the control sample that received the general instream flow question. The split-sample treatment for the group-size reminder directly preceded the valuation question and was written to closely follow that used in Green et al. (1994). For modeling, the presence of the group size reminder is hereafter indicated by the dummy variable, RM, where RM=1 indicates the reminder treatment, and RM=0 indicates no reminder.

The payment vehicle for the hypothetical market is a special trust fund used to buy or lease water from willing parties for the purpose of maintaining minimum instream flows. The trust fund payment vehicle was chosen to match those actually implemented by some states (e.g., Montana), and discussed in New Mexico. The voluntary contribution format is commonly used in CV studies of nonexclusive environmental goods, including the protection of instream flows (Duffield and Patterson, 1992; Brown and Duffield, 1995). Respondents are asked their willingness to contribute A($) annually for each of five years to protect minimum instream flows. The dichotomous choice valuation question for the treatment group (SM=1) is modified to identify minimum instream flows to specifically protect the silvery minnow in the Middle Rio Grande. Half of this treatment sample is crossed with the treatment for the group size reminder (RM=1)

An important element of the experimental design in dichotomous choice CV is the number and size of the offered payment amounts, A($), that are allocated across the sample. A large literature has developed around this topic with no clear consensus. The pragmatic approach chosen for the original 1995 survey was to iteratively choose 9 separate payment amounts to be allocated across the expected quarterly profile of 670 completed surveys. The identical bid structure, and as closely as possible its distribution across the sample, was then replicated in the FEB96 sample. In each case the set of nine separate payment amounts, A($)={5, 20, 30, 40, 50, 75, 100, 150, 200}, was also coordinated with the 2 x 2 experimental design to randomly allocate bids across the four treatment combinations.

The telephone survey included a wide variety of attitudinal and socio-economic questions. Descriptive statistics for selected variables for the FEB95 and FEB96 surveys are shown in Table 2.

THEORETICAL AND MODELING CONSIDERATIONS

For a posited change in instream flow from θ^0 (no minimum flow protection) to θ^1 (minimum flow protection), the household maximum willingness to pay, or Hicksian compensating variation (C), to acquire the increase in protection can be implicitly defined as:

$$U(P^0, \theta^0, Y^0) = U(P^0, \theta^1, Y-C) \tag{1}$$

Where U(•) is the utility function, P is a vector of prices for market goods, and Y is household income. The willingness-to-pay measure of the welfare change can be explicitly defined as:

$$WTP^C = E(P, \theta^1, U^0) - E(P, \theta^0, U^0). \tag{2}$$

where E(·) is the household's expenditure function and U^0 and initial level of utility (without instream flow protection). Thus, WTP^C is an income adjustment that represents the household's maximum willingness to pay to acquire the change in instream flow protection from θ^0 to θ^1 ($\theta^1 > \theta^0$), while maintaining utility at the initial level, U^0. It also implies that the property right is not currently held by those valuing instream flows, as is the case in New Mexico.

For the case of minimum instream flow protection, the protection outcome θ^1 can be viewed as vector of geographic components (rivers or river stretches) with individual elements θ_j^1. Then imposing the strong monotonicity condition on the valuation of any single geographical component (e.g., valuing protection on the Middle Rio Grande versus all major NM rivers) implies, $WTP^C(\theta_j^1) <$

$WTP^C(\theta^1)$. This is also a testable hypothesis (Carson and Mitchell, 1995) and provides the basis for our test of scope.

In practice, WTP is a stochastic variable and may be conditioned on a number of determinants. Further, in the dichotomous choice elicitation format, WTP is an unobservable variable and must be statistically inferred from the yes ($W_i=1$) and no ($W_i=0$) responses to the payment amount, A, which is varied across the sample. More specifically, let the probability of a yes response to the valuation question be

$$P(W_i = 1) = 1 - G\,[X, A;\, \beta,\, \kappa], \tag{3}$$

where G is an appropriate cumulative distribution function, X is a vector of covariates, or may simply be a scaler of ones, and β and κ are unknown location and scale parameters, respectively, of the distribution. Using information on the probability distribution of W_i across the bid levels, the parameters (β and κ) can be estimated using maximum likelihood techniques. The generic likelihood function is:

$$L = \Pi_{W_i=1}(P)\ \Pi_{W_i=0}(1-P)\ . \tag{4}$$

Conditional estimates of willingness to pay (WTP) are obtained when X contains covariates (e.g., socioeconomic characteristics) and β is a vector. The vector X may also include treatment indicators (e.g., RM, SM and FEB96) for hypothesis testing. A full review of statistical and theoretical modeling issues with dichotomous choice CV data is provided in Hanemann and Kanninen (1996).

Following the censored threshold approach to interpreting dichotomous choice data (Cameron, 1988), assume the individual has an underlying continuous linear WTP function over a vector of explanatory variables and an error term, $WTP_i = X_i'\beta + \upsilon_i$, where υ_i is a random variable, which in the logistic case has mean zero, standard deviation b, and scale parameter $\kappa = b\cdot(\sqrt{3})/\pi$. However, the individual's true WTP is an unobservable random variable whose magnitude must be inferred through a discrete indicator variable, W_i:

$$W_i = 1 \ if \ WTP_i \geq A_i; \ W_i = 0 \ otherwise \tag{5}$$

The bid amount, A_i, is thus a stimulus variable to which the individual reacts, accepting or rejecting if her true WTP is above or below this censoring threshold. The probability of answering yes is

$$\begin{aligned} P(W_i=1) &= Prob(WTP_i \geq A_i) \\ &= Prob(\nu_i \geq A_i - X_i'\beta) \\ &= Prob((\nu_i/\kappa) \geq (A_i - X_i'\beta)/\kappa) \ . \end{aligned} \tag{6}$$

Then (3) might be rewritten as:

$$P(W_i=1) = 1 - G(\frac{A - \mu}{\kappa}) \tag{7}$$

where $\mu = X_i'\beta$ in a linear regression context. Given the logistic assumption, for example, about the random variable υ_i/κ, $P(W_i = 1)$ takes the form

$$P(W_i=1) = [1 + e^{(A_i - X_i'\beta)/\kappa}]^{-1} \ . \tag{8}$$

Again, nonlinear optimization methods can be used to find maximum likelihood estimates of β and κ and their standard errors.

EMPIRICAL RESULTS

Test of Temporal Reliability

To test temporal reliability when replicating the survey instrument we begin by directly estimating separate WTP models (Cameron, 1988), under the logistic distributional assumption. Models 1 and 2 refer, respectively, to the FEB95 and FEB96 samples, while Model 3 uses the pooled sample. The evidence from a likelihood ratio (LR) test supports the null hypothesis of no significant difference in the vector of coefficients from the restricted case (pooled model) versus the unrestricted case (separate models).[5] The conclusion is that we can combine the FEB95 and FEB96 samples and that WTP functions are temporally stable.

Several caveats should be discussed. First, the results from the test of temporal reliability are presented here using only the logistic distributional assumption. However, the LR test result supporting the null hypothesis of no difference in the sets of coefficients for the FEB95 and FEB96 samples was found to be stable across a variety of alternative distributions. Second, in all specifications in Table 4, the change in scope is restricted to using the dummy variable SM, which is always negative and statistically significant, supporting the alternative hypothesis of sensitivity to the change in scope (H_0: $\beta_{SM} = 0$, versus H_A: $\beta_{SM} \neq 0$). However, using the FEB95 sample, Berrens et al. (1996) found statistical evidence that the vectors of coefficients differed across the change in scope (SM=1 and SM=0). For the test of scope below we present separate models for both levels of the good, and investigate four alternative distributional assumptions.

Test of Scope (Nested Geographical Components)

To test the sensitivity of valuation responses to the change in scope, Table 5 presents four sets of comparative results. Using the pooled data from both quarterly profiles, separate models are estimated for each level of the good (SM=0 and SM=1). Then this comparison is done using four different distributional assumptions: logistic, log-logistic, weibull and log-normal.

---

[5] The appropriate LR test statistic is: $q = -2(\ln L_{95+96} - (\ln L_{95} + \ln L_{96}))$, which has a χ^2 distribution, and where L_{95+96} refers to the log-likelihood value for Model 3 using the pooled data set. Using the results from Table 4, we get q=11.42 with a critical value of 18.55 at the 0.10 level (*df* =12).

The logistic distributional assumption allows predictions of both positive and negative WTP, but imposes symmetry (mean equal to median WTP). The commonly used log-logistic, log-normal and weibull distributions allow mean and median WTP to differ, but do not allow for negative values.

Before comparing estimates of conditional WTP across the change in scope ($WTP_{SM=0}$ versus $WTP_{SM=1}$), we first compare the separate β vectors of estimated coefficients. The testing strategy is to estimate separate WTP models for SM=0 and SM=1, and then compare against the pooled model. For each of the four distributional assumptions, the evidence from an LR test supports the alternative hypothesis of a significant difference at the 0.01 level in the vector of coefficients from the restricted case (pooled model) versus the unrestricted case (separate models).[6] The conclusion is that we cannot combine the SM=0 and SM=1 samples and that their WTP functions are not statistically equivalent.

Test of Information (Group-Size Reminder)

The test of information (sensitivity to the inclusion of a group-size reminder) is conducted using the modeling results presented previously in Tables 4 and 5. Consistent with the findings of Berrens et al. (1996) using the February 1995 quarterly profile sample (FEB95), we find no statistical evidence of sensitivity in valuation responses to the inclusion of the group-size reminder. In all specifications in Tables 4 and 5, the coefficient on the dummy indicator variable RM is never statistically significant; the evidence supports the null hypothesis (H_0: $\beta_{FEB96} = 0$ versus H_A: $\beta_{FEB96} \neq 0$). While not presented here, an interaction term between the group-size reminder and the change in scope (RM*SM) was also evaluated in various specifications and never found to be significant.

---

[6]The appropriate LR test statistic is: $q = -2(\ln L_{0+1} - (\ln L_0 + \ln L_1))$, which has a χ^2 distribution, and where L_{0+1} refers to the log-likelihood value for the estimated model using the pooled data set (SM=0 and SM=1). The pooled models are not presented in Table 5, but are available upon request. Using the results from Table 5, with a critical value of 26.22 at the 0.01 level (df =12), we get: q=34.63 for the logistic case, q=36.94 for the log-logistic case, q=36.60 for the weibull case, and q=36.16 for the log-normal case. In all cases the null that the WTP functions for the two goods are similar is rejected.

Conditional WTP Results

Using the modeling results from Table 5, a comparison of conditional WTP estimates, across the change in scope and using alternative distributional assumptions, is presented in Table 6. In addition to estimates of mean WTP, median WTP and the interquartile range (25th and 75th percentiles) are provided for both the SM=0 and SM=1 models. Additionally, both 90 and 95 percent confidence intervals (CI's), generated using the Krinsky-Robb approach of Park et al. (1991), are presented for every WTP estimate.

Estimates of mean WTP are extremely sensitive to the distributional assumption. For the logistic distribution, mean $WTP_{SM=0}$ is \$72 and mean $WTP_{SM=1}$ is \$26, with no overlap in the 95% CI's. Not uncommonly, mean WTP's are found to be infinity for the log-logistic case, given estimates of $\kappa > 1$. In the weibull distribution case, which is receiving increasing use in applied DC-CV studies, mean $WTP_{SM=0}$ is \$1819, mean $WTP_{SM=1}$ is \$206, and both the 90 and 95% confidence intervals overlap. In the log-normal distributional case, mean $WTP_{SM=0}$ is \$79,328, mean $WTP_{SM=1}$ is \$662, and both the 90 and 95% CI's overlap. Clearly, estimates of the mean are highly unstable in the absence of any truncation to the upper tail of the distribution. However, applying the convolutions approach (Poe et al., 1994) for comparing overlapping confidence intervals from simulated distributions, the evidence still supports the alternative hypothesis of a significant difference between the two estimates at the 0.0004 and 0.0001 levels, respectively, for the weibull and log-normal cases.[7]

Estimates of median WTP are much more stable than mean WTP, and in all cases lower than the corresponding estimate of mean WTP. Across all distributional assumptions median $WTP_{SM=0}$ is larger than median $WTP_{SM=1}$, with no overlap in any of the 90% and 95% CI's. For the log-logistic,

[7]A number of sensitivity checks on the convolutions results were also conducted, including changing the number of intervals used (e.g., 50, 100) in calculating the empirical CDF, and dropping outliers in the simulated series above \$10,000. In all cases the qualitative results remain unchanged.

weibull and log-normal cases median $WTP_{SM=0}$ is approximately \$55 and median $WTP_{SM=1}$ is approximately \$25. Thus, the evidence from the comparison of both mean and median WTP's supports an unambiguous conclusion--there is statistically significant sensitivity to the change in scope.

CONCLUSIONS AND FUTURE RESEARCH

Application of a carefully administered telephone survey instrument shows that NM households place a positive value on the protection of minimum instream flows. Berrens et al. (1996) argue that such values are important to the *prima facie* case--establishing sufficient evidence of the public benefits from maintaining instream flows to warrant consideration, or standing in future water policy deliberations (e.g., "beneficial use" or "public welfare" determinations). Further, confidence in the values estimated here is increased by the results of several spit-sample hypotheses tests.

First, the evidence supports sensitivity in valuation responses to a change in the scope of the good (nested geographical components). An important caveat with policy implications is that estimates of mean WTP are extremely sensitive to the distributional assumption, while estimates of median WTP are much more stable, and more conservative -- a not uncommon result.

Second, the evidence supports insensitivity in valuation responses to information concerning the collective nature of the provision of the public good (a group-size reminder). Since this is in stark contrast to the results of Green et al. (1994), an important future test will be a side-by-side comparison of the group-size reminder for open-ended and close-ended elicitation formats.

Third, the results from replicating the tests of scope (nested geographical components) and information (group-size reminder) demonstrate temporal reliability over a one year period. This reliability result is consistent with that of a number of recent CV studies (e.g., Carson et al., 1995).

Finally, consistent with Smith (1996), there is no evidence from this study to support the simple contributions framework for interpreting valuation responses.

REFERENCES

Arrow, K., R. Solow, E. Leamer, P. Portney, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register.* 58(1993):4601-14.

Berrens, R., P. Ganderton, and C. Silva."Valuing the Protection of Minimum Instream Flow Protection In New Mexico." *Journal of Agricultural and Resource Economics* 1996 (forthcoming).

Bestgen, K., and S. Platania. "Status and Conservation of the Rio Grande Silvery Minnow, Hybognathus Amarus." *The S.W. Naturalist.* 36(1991):225-32.

Bokum, C., V. Gabin, and P. Morgan. *Living Within Our Means: A Water Management Policy for New Mexico in the Twenty-First Century.* Santa Fe: NM Environmental Law Center, 1992.

Brown, T., and J. Duffield. "Testing Part-Whole Valuation Effects in Contingent Valuation of Instream Flow Protection." *Water Resour. Research.* 31(1995):2341-51.

Cameron, T. "A New Paradigm for Valuing Non-market Goods Using Referendum Format: Maximum Likelihood Estimation by Censored Logistic Regression." *J. of Envir. Econ. and Manage.* 15(1988):355-79.

Carson, R., W. Haneman, R. Kopp, J. Krosnick, R. Mitchell, S. Presser, P. Ruud, and V.K. Smith. "Temporal Reliability of Estimates from Contingent Valuation." Resources for the Future, Discussion Paper 95-37, 1995.

Carson, R., and R. Mitchell. "Sequencing and Nesting in Contingent Valuation Surveys." *J. of Envir. Econ. and Manage.* 28(1995):155-73.

Colby, B. "Enhancing Instream Flow Benefits in an Era of Water Marketing." *Wat. Resour. Research.* 26(1990):1113-20.

Colby, B. "Benefits, Costs and Water Acquisition Strategies: Economic Considerations in Instream Flow Protection." In: L. MacDonnell and T. Rice. (eds.) *Instream Flow Protection in the West.* Natural Resources Law Center, U. of Colorado, 1993.

Crawford, C., et al. *Middle Rio Grande Ecosystem: Bosque Biological Management Plan.* Albuquerque: U.S. Fish and Wildlife Service, 1993.

Cummings, R., P. Ganderton, and T. McGuckin. "Substitution Effects in CVM Values." *Amer. J. of Agr. Econ.* 76(1994):205-14.

DeYoung, T. "Protecting New Mexico's Instream Flows." In: L. MacDonnell, and T. Rice. (eds.) *Instream Flow Protection in the West.* Natural Resources Law Center, U. of Col., 1993.

Duffield, J., C. Neher, and T. Brown. "Recreation Benefits of Instream Flow: Application to Montana's Big Hole and Bitterroot Rivers." *Wat. Resour. Research.* 28(1992):2169-81.

Duffield, J., and D. Patterson."Field Testing Existence Values: Comparison of Hypothetical and Cash Transactions." In: *Benefits and Costs in Natural Resource Planning*. Western Regional Research Publication, W-133, 1992

Dumars, C., and M. Minis. "New Mexico Water Law: Determining Public Welfare Values in Water Rights Allocation." *Ariz. Law Rev.* 31(1989):817-39.

Green, D., D. Kahneman , and H. Kunreuther. "How the Scope and Method of Public Funding Affects Willingness to Pay for Public Goods." *Pub. Opin. Quart.* 58(1994):49-67.

Harpman, D., E. Sparling, and T. Waddle. "A Methodology for Quantifying and Valuing the Impacts of Flow Changes on a Fishery." *Wat. Resour. Research.* 29(1993):575-82.

Hanemann, W.M., and B. Kanninen. "The Statistical Analysis of Discrete-Response Data." In: I. Bateman and K. Willis. (eds.) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries*. Oxford: Oxford University Press, 1996 (forthcoming).

Kahneman, D., and J. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *J. of Envir. Econ. and Manage.* 22(1992):57-70.

Kahneman, D., and I. Ritov. "Determinants of Stated Willingness to Pay for Public Goods: A Study in the Headline Method." *J. of Risk and Uncertainty.* 9(1994):5-38.

Kahneman, D., I. Ritov, K. Jacowitz, and P. Grant. "Stated Willingness to Pay for Public Goods: A Psychological Perspective." *Psychological Science.* 4(1993):310-15.

Kanninen, B. "Bias in Discrete Response Contingent Valuation." *J. of Envir. Econ. and Manage.* 28(1995):114-26.

Loomis, J. "Measuring General Public Preservation Values for Forest Resources: Evidence from Contingent Valuation Surveys." In: W. Adamowicz, P. Boxall, M. Luckert, W. Phillips, and W. White. eds. *Forestry, Economics and the Environment.* Wallingford, UK: CAB International, 1996.

Loomis, J. "The Economic Value of Instream Flow: Methodology and Benefit Estimates of Optimum Flow." *J. of Envir. Manage.* 24(1987):169-79.

Loomis, J., and M. Creel. "Recreation Benefits of Increased Flows in California's San Joaquin and Stanislaw Rivers." *Rivers.* 3(1992):1-13.

McNally, M., and O. Matthews. "Changing the Balance in Western Water Law? Montana's Reservation System." *Nat. Resour. J.* 31(1995):671-93.

Park,T., J. Loomis, and M. Creel. 1991. "Confidence Intervals for Evaluating Benefits from Dichotomous Choice Contingent Valuation Data." *Land Economics* 67(1):64-73.

Poe, G., E. Severance-Lossin, and M. Welsh. "Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76 (1994):904-915.

Randall, A. "What Practicing Agricultural Economists Really Need to Know About Methodology." *American Journal of Agricultural Economics* 75 (1993):48-59.

Rinne, J., and S. Platania. "Fish Fauna." In: D. Finch and J. Tainter. (eds.) *Ecology, Diversity, and Sustainability of the Middle Rio Grande Basin.* Fort Collins CO: USFS, Rocky Mountain Forest and Range Experiment Station, 1995.

Sanders, L., R. Walsh and J. McKean. "Comparable Estimates of River Recreational Values." *Wat. Resour. Research.* 27(1990):1387-94.

Smith, V.K. 1996. Can Contingent Valuation Distinguish Economic Values for Different Public Goods? *Land Economics* 72(2):139-151.

U.S. Fish and Wildlife Service (USFWS). *The Potential Economic Consequences of Designating Critical Habitat for the Rio Grande Silvery Minnow.* Draft Report. New Mexico Ecological Services Field Office, 1996.

Ward, F. "Economics of Water Allocation to Instream Uses in A Fully Appropriated River Basin: Evidence From a New Mexico Wild River." *Wat. Resour. Research.* 23(1987):381-92.

Table 1. Disposition Table for the February 1995 and 1996 Quarterly Profile Telephone Surveys

	Local	Long Distance	Total
February 1995 (FEB95)			
Completed Interviews	357	369	726[a]
Failure To Contact (e.g., no answers, busy, and exceeded ten tries) [b]	224	184	398
Appointments Not Completed	58	43	101
Refusals	169	141	310
Language Barriers	14	21	35
February 1996 (FEB96)			
Completed Interviews	434	336	770[a]
Failure to Contact (e.g., no answers, busy, and exceeded ten tries) [b]	199	148	347
Appointments Not Completed	61	17	78
Refusals	166	166	332
Language Barriers	12	26	38

[a] Includes pre-test respondents, used for interviewer training and final wording changes, not included in the final survey data sets but used by the CATI system in tracking response rates. For the FEB95 survey there were 28 of these observations (726-28=698 observations). For the FEB96 survey there were 59 of these observations (770-59=711).

[b] Includes some call backs discontinued before the ten try limit when the target number of surveys was reached.

Table 2. Acceptance Rates by Payment Amount

Payment, A	*FEB95*	*FEB96*	*FEB95 + FEB96*
$5	53/83 (.64) [3]	47/67 (.71) [1]	100/150 (.67) [4]
$20	57/78 (.73) [4]	40/66 (.61) [4]	97/148 (.66) [8]
$30	40/97 (.41) [3]	52/91 (.57) [5]	92/188 (.49) [8]
$40	36/75 (.48) [5]	32/79 (.41) [1]	68/154 (.44) [4]
$50	27/74 (.37) [6]	37/93 (.40) [1]	64/167 (.38) [7]
$75	36/78 (.46) [2]	28/73 (.38) [5]	64/151 (.42) [7]
$100	24/82 (.29) [3]	21/76 (.28) [3]	45/158 (.29) [6]
$150	22/61 (.36) [2]	13/62 (.21) [2]	35/123 (.29) [4]
$200	5/40 (.13) [2]	17/52 (.33) [2]	22/92 (.24) [4]
Totals	300/668 (.45) [30]	287/659 (.44) [24]	587/1327 (.44) [54]

The numbers in parentheses are percentage rates; the bracketed [] numbers in selected cells give the number of unusable responses or failures to answer the valuation question; these observations are not used in calculating acceptance rates.

Table 3. Descriptive Statistics for Selected Variables.

Variable	Description	FEB95 (n=561) Mean	FEB95 (n=561) St.Error	FEB96 (n=564) Mean	FEB96 (n=564) St. Error
AGE	Age in years	42.52	14.94	43.14	15.38
IMPORT	Importance of instream flows: Scale 0-10; 0=Not important at all, 10=Extremely important.	8.22	2.02	8.57	1.86
RECOG	Should instream flows be recognized as beneficial use: 1=Yes, 0=No.	0.86	0.34	0.83	0.37
ENVIR-OR	Environmental organization member: 1=Yes, 0=No.	0.12	0.33	0.14	0.35
BERN-CO	Bernalillo County resident: 1=Yes, 0=No.	0.44	0.50	0.50	0.50
FISH-LIC	Own fishing license=Yes, 0=No.	0.45	0.50	0.44	0.50
POL-IDEO	Political ideology: Scale 1-7; 1=Strongly Liberal, 7=Strongly conservative.	4.33	1.53	4.31	1.54
INC	Household income categories 1-9, in $1000s: 1=(<$10); 2=($10-20); 3=($20-30); 4=($30-40); 5=($40-50); 6=($50-60); 7=($60-70); 8=($70-80); 9=(>$80).	4.14	2.28	4.60	2.62
INC1	Income categories 1-3	0.47	0.50	0.42	0.49
INC2	Income categories 4 and 5	0.28	0.45	0.26	0.44
INC3	Income categories 6-9	0.25	0.44	0.32	0.47
AWARE	Aware of New Mexico fish species on endangered list: 1=Yes, 0=No.	0.48	0.50	0.47	0.50
RM	Treatment for test of sensitivity to Reminder of group size: 1= received Reminder, 0= did not receive Reminder	0.50	0.50	0.52	0.50
SM	Treatment for test of sensitivity to scope of the good: 1=instream flows for silvery minnow, 0=instream flows for major NM rivers	0.51	0.50	0.49	0.50
INTERACT	Interaction term: SM*RM	0.25	0.43	0.28	0.45

Table 4. Estimation Results for WTP Models--Testing Temporal Stability.

Variables	Logit Model 1 n=561 Feb 1995	Logit Model 2 n=564 Feb 1996	Logit Model 3 n=1125 Feb 95 & 96
INTERCEPT	*-1.28 (-1.72)	**-1.60 (-2.14)	***-1.40 (-2.70)
AGE	**-1.27 (-2.03)	**-1.32 (-2.11)	***-1.27 (-2.89)
IMPORT	***2.34 (3.58)	***2.17 (3.11)	**2.16 (4.67)
RECOG	**0.83 (2.48)	***1.05 (2.99)	***0.95 (3.96)
POL-IDEO	*-1.09 (-1.81)	-0.83 (-1.34)	**-0.89 (-2.09)
INC2	**0.49 (2.11)	-0.06 (-0.27)	0.24 (1.54)
INC3	**0.78 (2.91)	0.07 (0.30)	**0.38 (2.35)
RM	0.13 (0.74)	0.27 (1.47)	0.19 (1.50)
ENVIRON-ORG	0.26 (0.92)	***0.81 (2.74)	***0.54 (2.73)
BERN-CO	0.11 (0.61)	*0.34 (1.81)	*0.21 (1.68)
SM	***-0.58 (-2.94)	**-0.41 (-2.16)	***-0.49 (-3.66)
κ (scale parameter)	***0.93 (5.30)	***0.95 (5.39)	***0.94 (7.63)
Log-likelihood	-333.46	-334.02	-673.19
LR Test (χ^2)	***82.95	***80.50	***149.09
McFadden R^2	0.11	0.11	0.10
AIC	1.2316	1.2270	1.2181

Numbers in parentheses are asymptotic t-statistics; *, **, *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. Standard errors were calculated using the heteroskedastic consistent variance-covariance matrix. To facilitate convergence in the nonlinear MLE optimization a number of variables were rescaled as follows: A/100, AGE/100, IMPORT/10, and POL-IDEO/10. The coefficient on κ is the negative of the inverse of the bid (A) coefficient in the direct logit model.

Table 5. Estimation Results for WTP Models--Testing Sensitivity to the Change in Scope

	Logistic		Log-Logistic		Weibull		Log-Normal	
Variables	SM=0	SM=1	SM=0	SM=1	SM=0	SM=1	SM=0	SM=1
INTER-CEPT	***-1.87	*-1.36	***-5.96	***-4.11	***-4.09	**-2.62	***-5.98	**-3.97
	(-2.89)	(-1.73)	(-4.03)	(-3.63)	(-3.25)	(-2.68)	(-.3.48)	(-3.55)
AGE	0.03	***-2.51	0.22	***-4.12	0.17	***-3.95	0.17	***-4.14
	(0.21)	(-3.55)	(0.19)	(-4.02)	(0.15)	(-3.90)	(0.07)	(-3.93)
IMPORT	***1.94	***2.55	***3.83	***4.10	***3.27	***3.74	***3,90	***4.04
	(3.44)	(3.34)	(3.20)	(3.99)	(3.27)	(4.05)	(3.24)	(3.93)
RECOG	***0.91	***0.90	***1.82	***1.54	***1.48	***1.36	***1,82	**1.53
	(3.00)	(2.60)	(2.88)	(3.05)	(2.89)	(2.96)	(2.83)	(3.01)
POL-IDEO	-0.49	**-1.22	-0.90	**-1.96	-0.69	**-2.25	-0.93	**-2.14
	(-0.87)	(-2.03)	(-0.77)	(-2.13)	(-0.65)	(-2.29)	(-0.75)	(-2.21)
INC2	0.18	0.29	0.41	0.49	0.18	0.49	0.38	0.51
	(0.85)	(1.37)	(0.94)	(1.49)	(0.48)	(1.39)	(0.86)	(1.49)
INC3	0.37	*0.37	0.74	*0.57	0.59	0.54	0.72	0.55
	(1.60)	(1.68)	(1.58)	(1.71)	(1.39)	(1.57)	(1.51)	(1.59)
RM	0.18	0.22	0.42	0.31	0.38	0.29	0.44	0.32
	(1.01)	(1.19)	(1.18)	(1.12)	(1.14)	(1.00)	(1.21)	(1.11)
ENVIRON-ORG	***0.94	0.18	***1.85	0.39	***1.92	0.31	***1.90	0.36
	(3.10)	(0.71)	(2.96)	(0.97)	(2.90)	(0.71)	(3.04)	(0.86)
BERN-CO	**0.41	0.02	**0.78	0.06	**0.69	0.15	**0.78	0.08
	(2.24)	(0.18)	(2.12)	(0.24)	(1.96)	(0.54)	(2.02)	(0.26)
FEB96	-0.28	-0.06	-0.54	-0.07	-0.48	-0.17	-0.55	-0.09
	(-1.57)	(-0.36)	(-1.46)	(-0.25)	(-1.40)	(-0.61)	(-1.46)	(-0.33)
κ (scale parameter)	***0.92	***0.91	***1.88	***1.40	***0.40	***0.44	***3.16	***2.42
	(5.91)	(5.08)	(5.14)	(6.38)	(5.19)	(6.73)	(5.22)	(6.66)
Log-likelihood	-335.98	-326.75	-338.29	-317.98	-338.82	-320.29	338.62	-318.90
LR Test (χ^2)	***78.30	***80.26	***86.00	***75.11	***73.03	***81.72	***74.60	***84.27
McFadden R^2	0.10	0.11	0.12	0.10	0.10	0.11	0.10	0.12
AIC	1.2340	1.2078	1.2422	1.1764	1.2440	1.1846	1.2433	1.1797
Minimum Distance	0.267	0.147	0.194	0.079	0.323	0.185	0.199	0.077

Numbers in parentheses are asymptotic t-statistics; *, **, *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. Standard errors were calculated using the heteroskedastic consistent variance-covariance matrix. To facilitate convergence in the nonlinear MLE optimization a number of variables were rescaled as follows: A/100, AGE/100, IMPORT/10, and POL-IDEO/10. Sample sizes are 564 for SM=0 and 561 for SM=1.

Table 6. Conditional WTP ($) Estimates Under Alternative Distributional Assumptions [a]

	Logistic		Log-Logistic		Weibull		Log-Normal	
	SM=0	SM=1	SM=0	SM=1	SM=0	SM=1	SM=0	SM=1
Mean WTP (st. err.)	72.18 (8.78)	26.42 (10.51)	*c*	*c*	1819.26 (31307)	206.13 (501.43)	79328 (377, 379)	662.14 (812.15)
90% CI	58 to 87	10 to 44			194 to 2840	95 to 399	546 to 311051	146 to 1891
95% CI	55 to 89	6 to 48			168 to 5168	87 to 517	372 to 592741	122 to 2522
Median WTP (st.err.)	*b*	*b*	52.13 (9.29)	24.11 (3.89)	57.74 (10.35)	26.46 (4.74)	52.22 (9.49)	23.87 (3.93)
90% CI			39 to 69	18 to 31	42 to 76	19 to 35	38 to 69	18 to 31
95% CI			36 to 73	17 to 33	40 to 80	18 to 36	36 to 73	17 to 33
25th prctl. (st.err.)	-29.62 (18.60)	-73.50 (25.17)	7.04 (3.04)	5.38 (1.88)	6.38 (2.61)	3.71 (1.51)	6.57 (2.86)	4.88 (1.76)
90% CI	-60 to 1	-115 to -32	3 to 13	3 to 9	2 to 11	1 to 6	3 to 12	3 to 8
95% CI	-66 to 7	-122 to -23	3 to 15	3 to 10	2 to 12	1 to 7	3 to 14	2 to 9
75th prctl. (st.err.)	173.98 (21.07)	126.34 (15.05)	449.19 (213.69)	113.81 (23.53)	379.23 (180.85)	133.21 (28.47)	484.89 (230.61)	123.30 (26.09)
90% CI	139 to 209	102 to 151	196 to 840	80 to 156	189 to 692	96 to 183	206 to 924	86 to 170
95% CI	133 to 215	97 to 156	171 to 960	75 to 166	172 to 821	91 to 200	180 to 1056	80 to 181

Notes:
[a] Means, medians and their confidence intervals are generated by 5000 repetitions of the Krinsky-Robb procedure (Park et al., 1991), with the upper tail of the distribution left untruncated.
[b] In the logistic model mean and median WTP are equal.
[c] In the log-logistic model the mean WTP and variance are undefined for the scale parameter $\kappa > 1$.

Appendix A: Telephone Survey Description and Selected Text

This appendix provides the text for the instream flow portion used in both the February 1995 and 1996 Quarterly Profiles. For brevity, the text for coding answers is not provided. All question answer codes included a Don't Know/No Answer (DK/NA) option. Each respondent was asked a set questions, renumbered below as Q1-Q5. To implement the 2 x 2 split sample treatments, the CATI system then directs survey observations through four possible paths for the contingent scenario and valuation questions Q6-Q11: (1). Q6, Q7, Q8; (2) Q6, Q8: (3) Q9,Q10, Q11; and (4) Q9, Q11. There were also several follow-up questions (e.g., open-ended WTP) to each dichotomous choice valuation question that are not replicated here.

Q1. The next series of questions concern water quality and water quantity in New Mexico. There are many competing demands for water found underground and in rivers, lakes and streams. These demands come from cities households, agriculture and industry.

How important do you think water issues are in New Mexico? Using a scale where zero is not at all important, ten is extremely important, and you may choose any number in between, please tell me how important you consider water issues in New Mexico?

Q2. Under New Mexico water law, water must be put to a beneficial use or the right to the water may be lost. Traditionally, beneficial uses include irrigated agriculture, industry and cities. Another possible use of water is to leave it in rivers and streams. Instream flow is a measure of the water in rivers and streams. Protecting instream flows ensures a certain amount of water flowing in rivers and remaining in lakes.

How important do you think it is to maintain minimum instream flows in the major rivers of New Mexico? Using a scale where zero is not at all important, ten is extremely important, and you may choose any number in between, please tell me how important you think it is to maintain minimum instream flows in the major rivers of New Mexico?

Q3. Instream flows support fish and wildlife, vegetation and habitat, recreation and viewing opportunities. Minimum instream flows can also protect water quality by diluting pollution. Maintaining instream flows may prevent costly federal government actions to protect endangered species and water quality.

At present New Mexico does not recognize instream as a beneficial use of water. If New Mexico were to recognize instream flows as a beneficial use, private individuals and groups, and government agencies could buy or lease water to be left in rivers and streams. It is possible that the price of some agricultural commodities and municipal water rates could increase, and some development could be restricted.

Do you think that instream flows should be legally recognized as a beneficial use of water?

Q4. In some states, government agencies such as Fish and Wildlife, or Parks and Recreation, can buy or lease water from willing parties in order to protect instream flows during low flow years.

Would you vote yes or no to allow a state agency to buy or lease water from willing parties in order to protect instream flows?

Q5. There are currently six fish species listed as endangered in New Mexico, with another five fish species listed as threatened.

Were you previously aware that any New Mexico fish species had been listed as endangered or threatened?

Q6. By federal law the critical habitat of endangered fish species must be protected, and this may require maintaining minimum instream flows. In New Mexico, endangered fish species are found in a number of the major rivers including the Gila, Pecos, Rio Grande and the San Juan.

The silvery minnow is a small fish found in the Middle Rio Grande and is currently listed as an endangered species.

Now I would like to ask you several questions about the dollar value your household puts on protecting minimum instream flows specifically to protect the silvery minnow.

There are no right or wrong answers. Before answering, remember your household income and budget, and decide what you could realistically afford. Money spent on protecting instream flows is money not available for other goods, public programs, or other environmental programs. The establishment of a special trust fund for buying or leasing water is used in some states to protect fish species.

Q7. If a special trust fund was set up in New Mexico, and requests were made statewide, up to half a million households could contribute. So, each dollar of average household contribution produces a half a million dollars for the special trust fund.

Q8. Would your household contribute $A dollars each year for five years to a special trust fund used to buy or lease water from willing parties to maintain minimum instream flows for the silvery minnow in the Middle Rio Grande?

Q9. By federal law the critical habitat of endangered fish species must be protected, and this may require maintaining minimum instream flows. In New Mexico, endangered fish species are found in a number of the major rivers including the Gila, Pecos, Rio Grande and the San Juan.

Now I would like to ask you several questions about the dollar value your household puts on protecting minimum instream flows.

There are no right or wrong answers. Before answering, remember your household income and budget, and decide what you could realistically afford. Money spent on protecting instream flows is money not available for other goods, public programs, or other environmental programs. The establishment of a special trust fund for buying or leasing water is used in some states to protect fish species.

Q10. If a special trust fund was set up in New Mexico, and requests were made statewide, up to half a million households could contribute. So, each dollar of average household contribution produces a half a million dollars for the special trust fund.

Q11. Would your household contribute $A dollars each year for five years to a special trust fund used to buy or lease water from willing parties to maintain minimum instream flows in the major rivers of New Mexico?

COMPARISON OF CV AND CONJOINT ANALYSIS IN GROUNDWATER VALUATION

Christopher Barrett, Thomas H. Stevens and Cleve E. Willis

ABSTRACT

CV and conjoint methods were used to value groundwater protection program alternatives. Conjoint estimates were much larger than the corresponding CV estimates; a result which is consistent with most previous comparisons for other commodities. However, the conjoint results compared favorably with averting expenditures, suggesting that conjoint may be a useful and credible alternative for valuing changes in environmental quality.

March 1996

INTRODUCTION

The contingent valuation technique (CV), which is frequently used to measure both use and nonuse values of environmental quality, is often viewed with skepticism[1] (Hausman, 1993). As a result, attention has begun to focus on modifications and alternatives to the traditional CV method, such as conjoint analysis which asks respondents to rate, rather than to price, alternatives. Although conjoint may have several potential advantages relative to the CV method, the validity and reliability of conjoint analyses for valuing nonmarket commodities is largely untested. This paper compares the CV and conjoint methods for valuing the benefits of environmental quality. A case study of the value of alternative groundwater protection programs is used to facilitate discussion.

BACKGROUND AND PREVIOUS RESEARCH

Some economists suggest that when compared to CV, conjoint analysis asks respondents to make decisions in a manner that is more familiar to them (McKenzie, 1990). The potential for hypothetical bias may therefore be reduced, and since conjoint respondents can express ambivalence or indifference directly, nonresponse and protest behavior may be reduced. Moreover, substitutes are made explicit in the conjoint format and this encourages respondents to explore their preferences and tradeoffs in detail. By focusing on the various attributes of commodities, each attribute can be valued separately, and the potential for embedding, wherein an individual's willingness-to-pay is not different for goods that differ with respect to scope or scale, may be minimized.

Empirical evidence presented by Ready, et al. (1991), Magat, Viscusi and Huber (1988) and Desvouges, et al. (1983) suggest that CV and conjoint results do, in fact, differ. Magat, Viscusi and Huber (1988) used a paired comparisons approach (PC), which is a form of conjoint analysis, and an open-ended CV format to derive consumers' willingness to pay (WTP) for risk reduction associated with a set of market goods (bleach and drain openers) that differ only in terms of purchase price and risk of injury. In contrast to the one-step open-ended CV elicitation procedure, the PC method asked each subject to make a series of comparisons between products in a manner which simulated actual choices in the marketplace. The direct CV approach produced monetary valuations that were 58% lower than the average conjoint valuation. Magat, et al. argue that;

> "...the CV approach may create incentives for respondents to state values which are somewhat below their true reservation prices for the commodities being valued, while the paired comparison approach eliminates these incentives to understate preferences, and thus it seems to provide more accurate measures of WTP." (p. 409)

Ready, et al. (1991) compared a dichotomous choice CV format to a polychotomous choice (PLC) format. Their CV format asked respondents to determine whether or not they preferred a given program while the PLC format gave six choices (i.e., definitely prefer, probably prefer, maybe prefer, maybe not prefer, probably not prefer, definitely not prefer). This format was motivated by the belief that respondents might be more comfortable answering valuation questions when given the opportunity to express strength of conviction; since PLC allows for a range of answers, it might produce a more accurate description of respondents' preferences.

In two empirical studies (preservation of wetlands and horse farms), the PLC format resulted in a higher rate of usable responses compared to the CV format (67% versus 60% for the wetland study, 58% versus 53.5% for horse farms). The polychotomous choice approach also resulted in much higher estimates of WTP for the two amenities.

Finally, Desvouges, et al. (1983) compared CV and contingent rankings for water quality in the Monogahela River. Mean water use values derived from direct CV questions were three to four times less than the values estimated from the contingent ranking approach. Thus, although very few comparisons have been published, conjoint analysis appears to have produced value estimates which are generally much higher than those derived from the traditional CV format.

THEORETICAL CONSIDERATIONS

From the perspective of neoclassical economic theory, CV and conjoint formats should produce similar results, provided that the conjoint format is properly specified. Following Viscusi, et al. (1991), individual utility associated with environmental factors related to human health, such as air or water quality, can be expressed as a function of income, Y, and health status, H or D, where D indicates sickness, and H, health. In dichotomous choice contingent valuation, the individual is asked to pay amount \$N for an environmental improvement or protection program. The expected value of utility, observed by the researcher, when amount N is paid is:

(1) $$EU_d = P_1\, U(D, Y\text{-}N) + (1\text{-}P_1)\, U(H, Y\text{-}N) + e_1$$

where P_1 is the probability, assigned by the individual, of environmental quality associated illness, and e_1 is a random variable. The expected value of utility when \$N is not paid is:

(2) $$EU_{Nd} = P_2\, U(D, Y) + (1\text{-}P_2)\, U(H, Y) + e_0$$

where $P_2 \geq P_1$. The individual is assumed to pay if, and only if:

(3) $$EU_d \geq EU_{Nd}$$

The willingness to pay probability can then be written as (Hanemann, 1984):

(4) $$P_r = G(dV)$$

where G is the probability function for the random component of utility and dV is the expected utility difference.

(5) $$dV = EU_d - EU_{Nd}$$

If utility is assumed to be additive and separable with respect to income and health status, dV is given by:

(6) $$dV = (P_1 - P_2)\,(U(D) - U(H)) + U(-N) + e_1 - e_0$$

Assuming a logit probability function for G, the willingness to pay probability is:

(7) $$P_r = (1 + e^{-dV})^{-1}$$

Median willingness-to-pay can then be estimated by calculating the value of N, N*, for which dV = 0, i.e., at the point of indifference there is a 50 percent chance that the individual would pay amount N*. Rearranging:

(8) $$N^* = \frac{(P_1 - P_2)\ (U(D) - U(H))}{b}$$

where b is the marginal utility of income. N* is therefore a function of the change in the probability of illness, the utility difference between the state of health and illness, and the marginal utility of income.

If a payment card CV format is used instead, each individual is assumed to select a value interval, N_1 - N_2, containing N*, for which dV = 0 such that from 6:

(9) $$N_1 - N_2 = \frac{(P_1 - P_2)\ (U(D) - U(H))\ +\ e_1 - e_0}{b}$$

where N* is contained within the N_1 - N_2 interval.

Results which are conceptually consistent with dichotomous choice CV (eq. 8) and the payment card approach (eq. 9) can also be derived from a conjoint format. Following Roe, et al. (1996) individuals are asked to rate the current situation given by (2) and a set of environ-mental quality protection programs including the program represented by (1). It is assumed that:

(10) $$R_1 = EU_d \text{ and } R_0 = EU_{Nd}$$

where R_1 and R_0 are individual ratings.

Utility difference, dV, is then given by the ratings difference, R_1 - R_0:

(11) $$dV = R_1\text{-}R_0 = (P_1\text{-}P_2)\ (U(D) - U(H)) + U(\text{-}N) + e_1 - e_0$$

The value of N* at the point of indifference is:

(12) $$N^* = \frac{(P_1 - P_2)\ (U(D) - U(H))\ +\ e_1 - e_0}{b}$$

For empirical comparisons of conjoint and contingent valuation, information about rating differences from the status quo as given by (11), the marginal utility of income, b, and proxies for the utility difference associated with health and illness are required. In the case study which follows an approximation of utility difference was estimated where:

(13) $$dV = R_1 - R_0 = \alpha_0 + b(-N) + c(T) + d(F) + e_1 - e_0$$

where N is the predetermined program cost, T represents different water quality protection programs which serve as a proxy for $(P_1 - P_2)$, and F is a set of tastes and preferences which provide a proxy for (U(D) - U(H)).

It is important to note that the conjoint specification presented in equations (11)-(13) differ from the traditional conjoint model which involves estimating the following relationship between ratings and program attributes:

(14) $$U_i = R_i = V(Z^K) + P_z = b_0 P_z + b_1 Z_1^1 + \ldots b_n Z_n^1 + e_i$$

where U_i is individual i's utility for an attribute bundle; R_i is the individual's rating, $U(\cdot)$ is the non-stochastic component of the utility function, Z^K is a vector of attribute levels, P_z is the price for the attribute bundle Z, and b is the marginal utility or weight associated with each attribute.

Setting the total differential of (14) to the point of indifference and solving:

(15) $$dU_i = b_0 dP_z + b_1 dZ_1^1 + \ldots = 0$$

yields marginal rates of substitution for the attributes Z_1^1. Since a price attribute, P_z, is included, the marginal utilities of all attributes can be rescaled into dollars and willingness to pay for each attribute may be derived:

(16) $$dP_z = -b_1 dZ_1^1 / b_0 \quad \text{or}$$

$$dP_z / dZ_1^1 = -b_1 / b_0$$

where (16) yields the marginal value of Z_1^1. Since this model assumes additive utility, the marginal WTP can then be aggregated to derive total WTP for a multi-attribute good. (Johnson, et al., 1995).

The conjoint model set forth in (13) differs from the traditional approach summarized in equations (14)-(16) in that the dependent variable in (13) is the ratings difference from the status quo and independent variables are changes in program attributes from the status quo. As shown by Roe, et al. (1996), this specification provides estimates of Hicksian surplus, as opposed to marginal values of attributes, which can then be directly compared with CV results (also see McKenzie, 1990 and Johnson, et al., 1995).

Another important aspect of the ratings difference model is that in the traditional specification different respondents tend to center on different ranges of the rating scale. Roe, et al. argue that ..."using the status-quo rating as a common anchoring point for constructuring the rating difference helps remove this noise from the data".

CASE STUDY

To facilitate comparison of the CV and conjoint formats, two groundwater valuation surveys were administered to randomly selected residents of 56 Western Massachusetts towns containing a mix of suburban and rural communities which rely primarily on groundwater. The first survey used the contingent valuation technique and was mailed to 997 households in 1994 (Krug, 1994). The second survey employed a conjoint format and was mailed to 1054 households in 1995. Although both surveys targeted the same geographical area, respondents to the 1994 CV survey were not resurveyed in 1995; rather two independent samples were drawn from the same region.

Dillman's (1978) total design method was employed and focus groups were used to develop and pretest both surveys. Table 1 compares socio-economic characteristics of CV and conjoint respondents to each other and to nonmetropolitan Massachusetts residents as a whole. As shown in Table 1, CV and conjoint survey respondents were quite similar in all respects, but average age, education, and gender of respondents were significantly different from that of the average Massachusetts resident.[2]

Both surveys asked about each household's source of water, averting behavior and level of knowledge about groundwater.[3] Results are presented in Table 2 which, as expected, shows relatively little difference between CV and conjoint respondents.

The CV questionnaire asked for WTP for two types of groundwater quality programs; a town-wide aquifer protection district which would protect the resource itself, and a private pollution control device. One-half of the CV sample received the aquifer protection district question; all respondents were asked for their perception about the current level of safety of their groundwater supply and for their WTP to "insure that their groundwater quality does not get any worse". That is, ex ante and ex post (after investment in protection program) health risks were not objectively specified. Rather, as indicated in equations (1)-(12), individual perceptions of safety were implicit in both the CV and conjoint formats. CV questions for both the aquifer protection district and the private control device are presented in Appendix A (Krug, 1994).

The conjoint format presented respondents with background information about five options; an aquifer protection district, town-wide water treatment facility, private pollution control device, purchase of bottled water, and do nothing (status quo). Appendix B presents the information provided to respondents about these program options. As shown in Appendix B, method of protection, cost, length of payment, and participation were the key attributes which comprised the protection program options. The range of cost values associated with these options corresponded to the median values presented to CV respondents. Time spans of five and ten years were chosen to test whether differences in length of payment affected program ratings. There were four protection options, 14 price levels, two levels of participation and two payment schedules which make 224 possible scenario permutations. Since each attribute does not have the same number of levels or alternatives, the conjoint question design is asymmetric. Use of a fractional-factorial design resulted in only 112 different combinations for consideration because some attributes were incompatible with each other, or were not realistic. To generate the protection program options used in the survey, the 112 different combinations were generated by computer and random picks were then taken from this list four times. The random choices and a status-quo option comprised the conjoint question.[4] To ensure sufficient variability, sixty random scenarios were created using the methods described above. These were then duplicated eighteen times for a total of 1080; ultimately 1054 were sent to western Massachusetts residents.

Four program options and the status quo were rated by each respondent on a scale of one to ten, with ten indicating that the respondent would definitely vote in favor of the program, and one indicating that the respondent would definitely <u>not</u> vote for the program. If respondents were not sure they were asked to use a scale of 2 through 9 to indicate how likely they would be to vote for the options presented. An example of the conjoint question format is presented in Appendix B. After respondents completed the conjoint question, an open-ended question asked them to think about the factors considered in deciding about program ratings.

In evaluating and interpreting the results which follow, it is important to note that the CV and conjoint survey questions were not identical in all respects. The conjoint format provides more information about substitutes and commodity attributes and there was a one year difference between the CV and conjoint applications. Neither survey identified the current groundwater condition, the consequences of groundwater contamination were not specified, and changes in consequences were not defined in detail; the resulting value estimates therefore refer to protection of water quality in general, as perceived by each respondent, and are not necessarily comparable with previous studies of specific contaminants, such as nitrates. However, since the CV and conjoint surveys were identical in this respect, comparison of these techniques should not be affected.

RESULTS

Response rates associated with the CV and conjoint surveys were 44.7 and 51.0 percent respectively; as hypothesized the conjoint survey performed better in this respect. However, item nonresponse was slightly higher in the conjoint questionnaire; 24.8% of respondents did not answer the ratings question while 21.2% either did not respond or gave a protest zero bid to the CV question.

CV value estimates were taken directly from the questionnaire (see Q-19, Appendix A)[5] and conjoint value estimates were derived by regressing rating differences, as defined in (13), against program attributes and several independent variables representing individual tastes and preferences as defined in Table 3. A traditional conjoint model with the dependent variable expressed in terms of ratings as opposed to rating differences was also estimated for comparison (McKenzie, 1993).

Since the dependent variable in conjoint analysis takes on discrete values, such as integers from 1 to 10, an ordinary least squares estimating procedure is inappropriate. Two estimating techniques, ordered logit and doubly censored tobit, were applied to the ratings data. The ordered logistic model treats the dependent variable as an ordinal ranking of preferences while the doubly-censored tobit model assumes that the dependent variable is a cardinal measure (Roe, et al., 1996). When conjoint models are estimated in the ordered logistic form, the intercept term is decomposed into k-1 separate dummies to account for the intervals between rating levels as follows (McKenzie, 1990):

$$\begin{aligned}\text{Rate diff or Ratings} = \text{Alpha1} \ldots \text{Alpha9} &+ b_0\,\text{Price} + b_1\,\text{Aquifer} + b_2\,\text{Plant} + b_3\,\text{Filter}\\ &+ b_4\,\text{Bottled} + b_5\,\text{Length} + b_6\,\text{Avert} + b_7\,\text{Rate} + b_8\,\text{Info1}\\ &+ b_9\,\text{Info2} + b_{10}\,\text{Info3} + b_{11}\,\text{Own} + b_{12}\,\text{Gender} + b_{13}\,\text{Age}\\ &+ b_{14}\,\text{Educ} + b_{15}\,\text{Income} + e\end{aligned}$$

As shown in Table 4, estimated coefficients of the traditional ratings model were generally of the expected sign and magnitude. For example, coefficients for price and program type (Aquifer, Plant, Filter, and Bottled) variables, which are essential for calculation of the value of protection programs, were all statistically significant. As expected, rating declined with price and increased with averting behavior.

As shown in (16), marginal values can be derived from the traditional conjoint model (Table 4) for each groundwater protection program by taking the negative of the ratio of the estimated coefficient for each attribute divided by the estimated price coefficient. For example, taking the necessary coefficients from the ordered logistic regression (see Table 4), the marginal economic value of the four types of protection programs are calculated as follows:

$$MV_{\text{Aquifer Protection District}} = -b_1/b_0 = -[(1.8376)/(-0.00509)] = \$361.02$$

$$MV_{\text{Water Treatment Plant}} = -b_2/b_0 = -[(1.0498)/(-0.00509)] = \$206.25$$

$$MV_{\text{Private Filter}} = -b_3/b_0 = -[(1.6165)/(-0.00509)] = \$317.58$$

$$MV_{\text{Bottled Water}} = -b_4/b_0 = -[(0.3907)/(-0.00509)] = \$76.76$$

Results derived from the rating difference model are presented in Table 5. The ordered logistic estimation technique did not produce maximum likelihood estimates for this model because of a quasi-complete separation of the sample points. Consequently, only Tobit results are reported in Table 5. Compared with the traditional ratings model, the rating difference specification yielded more significant variables; individuals who rated their current water quality highly gave protection programs a lower rating relative to the status quo. Also, individuals who said they were very well informed about water quality gave lower rating differences. Protection program rating differences from the status quo also declined with age and education.

Value estimates derived from the two conjoint models are compared to the corresponding CV estimates in Table 6.[6] Confidence intervals were produced by bootstrapping and although there is some overlap with the CV estimates, conjoint estimates for these two types of protection programs were generally much larger than CV estimates, a result which is consistent with most previous comparisons for other commodities. Also, when compared to CV, the conjoint estimates were much more precise. On the other hand, the value of the aquifer protection program, which presumably includes both use and nonuse values, was not statistically different from the value of the private water filter option.[7]

Our principle concern, however, is with the disparity between the CV and conjoint value estimates. Several factors may be responsible for this difference. Conjoint responses are expressed in terms of ratings which may contain both ordinal and cardinal information (Roe, et al., 1996). Also, the conjoint format generally provides more information about substitutes and commodity attributes. On the other hand, ratings do not indicate whether respondents are actually in the market for the commodity being valued which may bias conjoint results upward. Also, compatibility and prominence effects may differ between the CV and conjoint formats. For example, Irwin, et al., (1993) argue that,

> "The compatibility effect implies that when dollars are an available (recognizable) attribute of an object, they carry more weight or influence in determining an equivalent response that is also in dollars (e.g., cash equivalent, selling price) than they do in determining a response that is not in dollars (e.g., a rating of value or a choice). The prominence effect causes choice responses to be more dominated by prominent attributes than are pricing responses. This arises from the fact that choices are driven by reason and arguments to a greater extent than are pricing responses" (p. 6, 7).

Although the relative accuracy of CV and conjoint measures cannot be proven, comparison with values derived from previous studies and from alternative methods provide benchmarks for evaluating both the CV and conjoint results reported in this study.

Results reported in previous groundwater CV studies are summarized in Table 7. The CV results derived in the present study are generally quite similar to those reported for the Northeast (Powell, 1991; Schultz and Lindsay, 1990). Although Edwards (1988) found values which were considerably larger, water is much scarcer on Cape Cod and water quality problems are much more common there. We are aware of only one previous conjoint study of groundwater quality (Sparco, 1995). Sparco focused on potential health risks associated with nitrate, atrazine and coliform contamination in Sussex County, Delaware. His model yielded an annual WTP estimate of $124 per household for a one part per million decrease in nitrate contamination; a result which appears to be consistent with several of the previous CV studies summarized in Table 7. However, since Sparco's results represent marginal values, from a conceptual perspective, they are not necessarily directly comparable to CV values.

Following Bartik (1988), Abdalla (1990) and Abdalla, Roach and Epp (1992), the averting expenditure method provides an alternative estimate of willingness-to-pay for groundwater quality. It is important to note, however, that actual changes in averting expenditures underestimate the theoretical lower bound of the benefits derived from non-marginal changes in water quality, and that "the cost of the least cost technology providing complete protection gives an upper bound estimate of benefits of the environment improvement" (Abdalla, p. 455).[8]

Despite these problems, evidence of averting expenditures and the cost of the least cost technology providing complete control place both CV and conjoint estimates into perspective. Following the approach employed by Abdalla, a survey of water treatment suppliers in western Massachusetts yielded an annual average cost of $322.81 for a point of entry granular activated carbon filtration system. This compares favorably with Abdalla's average annual cost estimate of $382.91, which as noted by Bartik (1988) should be viewed as an upper bound estimate of the use value associated with groundwater quality. On the other hand, the lower bound of use value can be approximated by actual averting expenditures which were estimated by Abdalla to average $252 per household per year.[9]

Although these comparisons must be used with caution, it is interesting to note that results from averting expenditure and least cost treatment analyses bound those derived from the conjoint approach used in the study (see Table 6).

CONCLUSIONS

Independent samples of CV and conjoint surveys were used to value groundwater protection program alternatives. Conjoint estimates for two types of programs, a private filtration system and a townwide aquifer protection district, were found to be much larger than the corresponding CV estimates; a result which is consistent with most previous CV/conjoint comparisons. This disparity is cause for concern about the validity of survey techniques for valuing groundwater. However, conjoint estimates compared favorably with the actual market cost of a water filtration system, and compared with CV, the conjoint estimates were much more precise, suggesting that conjoint may be a useful and credible alternative for valuing nonmarket environmental commodities such as groundwater quality.

REFERENCES

Abdalla, C. W. "Measuring Economic Losses from Groundwater Contamination: An Investigation of Household Avoidance Costs". Water Resources Bulletin 26(1990):451-463.

Abdalla, C. W., B. Roach and D. Epp. "Valuing Environmental Quality Changes Using Averting Expenditures: An Application to Groundwater Contamination". Land Economics 68 No. 2 (1992):163-69.

Bartik, T. J. "Evaluating the Benefits of Nonmarginal Reductions in Pollution Using Defensive Expenditures". Journal of Environmental Economics and Management 15(1988):111-127.

Bergstrom, J. C. and K. J. Boyle. Using Benefits Transfer To Value Groundwater Benefits: Conceptual Issues. Paper presented at AAEA/AERE Annual Meeting, Orlando, Florida, 1993.

Desvousges, William H., and V. Kerry Smith. "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monogahela River". Journal of Environmental Economics and Management 14(1983):248-267.

Dillman, Don A. Mail and Telephone Surveys: The Total Design Method New York: John Wiley & Sons, 1978.

Edwards, S. F. "Option Prices for Groundwater Protection". Journal of Environmental Economics and Management 15 No. 4 (1988):475-487.

Hanemann, W. M. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." American Journal of Agricultural Economics 66(1984):322-341.

Hausman, J. A. (ed.). Contingent Valuation: A Critical Assessment. North Holland, New York, N.Y, 1993.

Irwin, J. R., et al. "Preference Reversals and The Measurement of Environmental Values". Journal of Risk and Uncertainty. 6(1993):5-18.

Johnson, F. Read, William H. Desvousges, Erin E. Fries, and Lisa L. Wood. Conjoint Analysis of Individual and Aggregate Environmental Preferences. Technical Working Paper No. T-9502, Triangle Economic Research, 1995.

Krug, Donald C. An Application of the Contingent Valuation Method to Groundwater Quality in Western Massachusetts. Master's Thesis. Department of Resource Economics, University of Massachusetts at Amherst, 1994.

McKenzie, John. "Conjoint Analysis of Deer Hunting." Northeastern Journal of Agricultural and Resource Economics 19 No. 2 (1990):109-117.

McKenzie, John. "A Comparison of Contingent Preference Models." American Journal of Agricultural Economics. 75(1993):593-603.

Magat, Wesley A., W. Kip Viscui and Joel Huber. "Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation." Journal of Environmental Economics and Management 15(1988):395-411.

Poe, G. L. Information, Risk Perceptions and Contingent Values: The Case of Nitrates in Groundwater. Ph.D. Dissertation. Department of Agri. Economics, University of Wisconsin, Madison, 1993.

Powell, J. R. The Value of Groundwater Protection: Measurement of Willingness To Pay Information, and Its Utilization By Local Government Decision-Makers. Ph.D. Dissertation. Cornell University, 1991.

Ready, R., J. Whitehead and G. Blomquist. "Polychotomous Choice Valuation Questions" W-133 Benefits and Costs in Natural Resource Planning, Fourth Interim Report, Department of Agricultural Economics, University of California, Davis, 1991.

Roe, Brian, Kevin J. Boyle and Mario Teisl. "Deriving Estimates of Compensating Variation From Conjoint Data." Journal of Environmental Economics and Management, forthcoming, 1996.

Schultz, S. and B. Lindsay. "The Willingness To Pay for Groundwater Protection". Water Resources Research 26 No. 9 (1990):1869-1875.

Sparco, J. "A Conjoint Analyses of Willingness To Pay For Changes in Groundwater Quality". Paper presented at NAREA Annual Meetings, Burlington, VT, 1995.

Viscusi, W. Kip, W. Magat and Joel Huber. "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Tradeoffs for Chronic Bronchitis." Journal of Environmental Economics and Management 21(1991):32-51.

Table 1 Socio-Economic Characteristics of Respondents[1]

Characteristic	CV Survey	Conjoint Survey	Census (1990)
Residence - Primary	95.2	94.9	88.5
Weekend/Vacation	4.8	5.1	11.5
Own Home	84	92	67.4
Years at residence	16.9	17.7	11
Age (Years)	51.5	51.9	35.0
Gender (% Male)	63	69	48
Education -			
Grade School	5	4	19
High School	37	35	57
College	38	39	15
Graduate School	20	22	9
Median Household Income (in $)	40,000	44,318	42,133[2]

[1] All values are percents unless otherwise noted.

[2] Derived by using the 1989 nonmetropolitan Massachusetts median income of $31,440 and an interest rate of 5%.

Table 2 Summary Statistics of Water Knowledge Questions[1]

Question	CV Survey	Conjoint Survey
Source of Water -		
Private	48	52
Public	49	46
Other	3	2
Water Utilities Source -		
Groundwater	47	46
Surface water	17	15
Combination	13	18
Other, or Didn't Know	23	21
Averting Behavior -		
Installed water filter	23	13
Drilled new well	9.2	4
Boiled tap water	19.5	7
Bought bottled water	34.7	18
Respondents Who Had Water Tested -	40	34
Respondent's Level of Water Knowledge -		
Very well informed	21	16
Know something	35	37
Know little or nothing	44	47[2]

[1] All values are percent unless otherwise noted.

[2] This figure is comprised of 27% of respondents who knew little about their water, and 20% of respondents who knew nothing about their water quality.

Table 3 Variables Used in Econometric Analysis of the Conjoint Data

Variable	Definition
Rate diff	Elicited rating of proposed protection program minus the rating of the status quo
Ratings	Elicited rating of proposed protection program
$Alpha_k$	Rating interval dummies (k=10)
$Alpha_i$	1 if the rating is i, and 0 otherwise
Aquifer	1 if groundwater protection program, 0 otherwise
Price	14 cost levels within the range of $0 to $325
Plant	1 if water treatment plant, 0 otherwise
Filter	1 if private water filter, 0 otherwise
Bottled	1 if bottled water program, 0 otherwise
Length	1 if length of payment is 10 years, 0 if payment is for 5 years
Avert	1 if respondent engaged in averting behavior, 0 otherwise
Rate	1 if respondent rated their water quality greater than 6, 0 otherwise
Info1	1 if respondent was very well informed about the quality of their groundwater, 0 otherwise
Info2	1 if respondent knew something about the quality of their groundwater, 0 otherwise
Info3	1 if the respondent knew little about the quality of their groundwater, 0 otherwise
Own	1 if respondent owned their home, 0 if rents
Gender	1 if male, 0 if female
Age	Respondents' age
Educ	Respondents' education level in years completed. 8 = Grade school, 12 = High school, 16 = College, 18 = Graduate school
Income	Respondents' income level. Median value of respondents' answer to question Q-21 in the survey (i.e., 1 = $5,000; 2 = 14,999.5; 3 = 29,999.5; 4 = 49,999.5; 5 = 69,999.5; 6 = 80,000)

Table 4 Traditional Conjoint Regression Results for the Ratings Model[1]

	ORDERED LOGISTIC PROCEDURE		TOBIT PROCEDURE	
Variable	Coefficient	Standard Error	Coefficient	Standard Error
Alpha1	-0.1914	0.4033	1.2661	1.295
Alpha2	0.3025	0.4034		
Alpha3	0.5716	0.4036		
Alpha4	0.7508*	0.4038		
Alpha5	1.3024**	0.4047		
Alpha6	1.4749**	0.4051		
Alpha7	1.7311**	0.4057		
Alpha8	2.2116**	0.4072		
Alpha9	2.7192**	0.4095		
Price	-0.00509**	0.00052	-0.0159**	0.0016
Aquifer	1.8376**	0.1664	5.4172**	0.5231
Plant	1.0498**	0.1719	3.0493**	0.5417
Filter	1.6165**	0.1781	5.0502**	0.5618
Bottled	0.3907**	0.1747	1.2044**	0.5562
Length	-0.0567	0.1045	-0.2756	0.3322
Avert	0.2299**	0.1008	0.7055**	0.3202
Rate	-0.0090	0.1303	-0.2464	0.4331
Info1	-0.2130	0.1711	-0.5921	0.5577
Info2	0.0215	0.1368	0.1317	0.4477
Info3	-0.0169	0.1428	0.1624	0.4688
Own	-0.2410	0.1834	-0.6671	0.5739
Gender	0.00459	0.1083	0.02849	0.3472
Age	-0.00429	0.000359	-0.0065	0.0114
Educ	0.0314	0.0206	0.11473*	0.0656
Income	2.78E-6	2.496E-6	0.101E-4	0.7E-5
σ			5.2029**	0.1490
'N'	1510		1434	
Amemiya's Pseudo R^2 = 0.042			-2 Log likelihood = 5962.136	

[1] *Indicates significance at the 10% level, ** Indicates significance at the 5% level.

Table 5 Regression Results for the Ratings Difference Model [1]

TOBIT PROCEDURE

Variable	Coefficient	Standard Error
Constant	4.2541**	1.103
Price	-0.0111**	0.00138
Aquifer	3.7472**	0.4450
Plant	2.2313**	0.4626
Filter	3.4228**	0.4787
Bottled	1.0827**	0.4707
Length	-0.38859	0.2857
Avert	1.7225**	0.2755
Rate	-0.92012**	0.3728
Info1	-2.3100**	0.4701
Info2	-0.35618	0.3813
Info3	-0.13096	0.4000
Own	-0.48659	0.4956
Gender	0.25829	0.2974
Age	-0.05050**	0.0098
Educ	-0.03834*	0.0556
Income	-0.470E-5	0.68E-5
σ	4.7211**	0.09709
'N'	1434	

-2 Log likelihood = 7942.516

[1] *Indicates significance at the 10% level, ** Indicates significance at the 5% level

Table 6 Comparison of CV and Conjoint WTP Estimates*

Type of Protection Program	CV Estimate (Average $/household/year)	Ratings Estimate (Average $/household/year)	Rate Diff Estimate (Average $/household/year)
Aquifer Protection District	$63.12 (85.50)	$361.02 (279 - 405)	$302.58 (220 - 373)
Private Water Filter	$74.67 (95.65)	$317.58 (242 - 365)	$273.35 (238 - 396)

*Figures in parentheses are standard error and 95% confidence intervals for CV and conjoint estimates, respectively.

Table 7 Summary of Previous Groundwater Contingent Valuation Studies[1]

Study	Valuation Issue	Survey Site	Response Rate (%)	Returned Surveys ('n')	Mean WTP[2]
Edwards (1988)	Protection of groundwater from nitrates	Cape Cod, MA	58.5	585	\$363 - 1,437
Schultz and Lindsay (1990)	Protection of groundwater from general contamination	Dover, NH	59.3	591	\$129
Powell (1991)	Groundwater resources in the Northeast	Massachusetts New York Pennsylvania	49.7	1,041	\$55.79 - 81.86
Sun (1992)	Protection of groundwater from nitrates and pesticides	Dougherty Co., GA	47.8	346	\$641
Caudill (1992)	Analysis of benefits of groundwater protection	Michigan	65.7	1,213	\$45.07 - 64.52
McClelland, et al. (1993)	Protection of groundwater contaminated by landfills	United States	60	3,000	\$146.76
Jordan and Elnagheeb (1993)	Protection of groundwater from nitrates	Georgia	35	192	\$120.84 - 148.56
Poe (1993)	Protection of groundwater from nitrates	Wisconsin	55	339	\$224.72 - 684.95

[1] Table adapted from Bergstrom and Boyle (1992).

[2] Mean WTP per household per year, unless otherwise noted.

APPENDIX A
AQUIFER PROTECTION DISTRICT

One way to prevent groundwater pollution is to establish a town-wide special aquifer protection district. This district would develop and implement pollution prevention programs specifically designed to suit the needs of your town. Households on public water systems and those with private wells would all benefit. Examples of possible actions include drilling new wells in an area where the water is uncontaminated and development restrictions on land near well fields and in aquifer recharge areas. Assume that these programs would be paid for by an increase in water utility bills for those on public water supplies, and by an increase in property taxes for households served by private wells. These programs would benefit your household, everyone else in your town, as well as future generations.

Q-19. Of the amounts listed below, how much would you be willing to pay **per year** for groundwater pollution prevention programs that insure that the quality of groundwater in your town does not get any worse? That is, to keep it at the same level of safety that you selected in question 17 above? **Circle one dollar range as your annual payment.**

$0	$1-10	$11-20	$21-30	$31-40
$41-50	$51-75	$76-100	$101-125	$126-150
$151-200	$201-250	$251-300	$301-351 or more	

If you would be willing to pay more than $350, what is the maximum amount that you would pay per year: $_________

PRIVATE CONTROL DEVICE

One way to prevent pollution of your water supply is to install pollution control devices, such as granular-activated carbon filters, on your tap. Potentially harmful particles and chemicals would be removed from your tap water; the quality of the groundwater itself would not be protected, however. Others outside your household and future generations would therefore not benefit.

Q-19. Of the amounts listed below, how much would you be willing to pay **per year** for a pollution control device on your tap that would insure that your water quality does not get any worse? That is, to keep it at the same level of safety that you chose in question 17 above? Please assume that no other programs will be undertaken to protect groundwater.

Circle one dollar range as your annual payment.

$0	$1-10	$11-20	$21-30	$31-40
$41-50	$51-75	$76-100	$101-125	$126-150
$151-200	$201-250	$251-300	$301-351 or more	

If you would be willing to pay more than $350, what is the maximum amount that you would pay per year: $_________

APPENDIX B
Conjoint Survey Information; Types of Groundwater Protection Programs

On the next page, you will be asked to rank five different types of groundwater protection programs. The following is a list of program features such as cost and method of protection. Please consider the following information very carefully, even if the groundwater supply for your household or community is already protected.

I. Method of Protection

(You will be asked for your opinion about one or more of the following protection methods.)

- Creation of a **town-wide groundwater protection district**. Examples of possible actions include drilling new wells in areas where the water is uncontaminated, and placing development restrictions on land near well fields or in groundwater recharge areas.
- Construction of **water treatment facility**, which would either filter or chemically treat the water as it is pumped from the ground.
- Installation of **private pollution control device**. Households could install a device such as a granular-activated carbon filter on their taps. These filters would clean all water delivered to the household by either the public water utility or the household's own private well.
- Purchase of **bottled water**. Households could purchase bottled water to substitute for the groundwater that they receive from either a public water utility or the household's private well.

II. Cost of Program

- Depending on the method of protection and other factors, the cost of groundwater protection is estimated to range from **$0** to **$325** dollars per household per year.

III. Length of Payment

- You may be asked to consider two alternatives; a payment each year for the next **5** years, or a payment each year for the next **10** years.

IV Who Would Benefit

- Some programs benefit the **participating household** only, others benefit **all households on public water**, and some benefit **all households**. Some programs may benefit the **present generation** only, others have the potential to benefit **future generations** as well.

V. Participation

- Groundwater protection programs may either be **voluntary** or **mandatory**. Mandatory programs must be passed by a majority of voters in the community.

Q-12. Now consider the five groundwater protection programs presented below. Please indicate how you would rank these programs on a scale of 1 to 10. Use 10 for the program, if any, you would definitely vote in favor of, and 1 for the program that you would definitely not vote in favor of. If you are not sure, use 2 thru 9 to indicate how likely you would be to vote for the options presented.

OPTION A

- Construction of a water treatment plant.
- Present and future residents on public water will benefit.
- The program would be paid for by an increase in your household's water utility bills of $325 per year for the next 5 years.
- Program mandatory if passed by a majority of voters in your town.

RANK: ________

OPTION B

- Town-wide groundwater protection district.
- All residents, both present and future, will benefit.
- The program would be paid for by an increase in your household's water utility bills or property taxes of $138 per year for the next 5 years.
- Groundwater quality would not get any worse.
- Program mandatory if passed by a majority of voters in your town.

RANK: ________

OPTION C

- No new groundwater protection program will be implemented.
- Maintain current level of protection of groundwater in your community.
- Water quality may decline over time due to economic growth and development.
- No increases in costs to you for groundwater protection.

RANK: ________

OPTION D

- Install private pollution protection device on water tap.
- Only households which participate will benefit.
- Installation, operation, and maintenance of the pollution protection device will cost your household $275 per year for the next 10 years.
- This program is voluntary.

RANK: ________

OPTION E

- Purchase bottled water.
- Only households which participate will benefit.
- Assume that bottled water to meet your household's needs will cost your household $138 per year for the next 5 years.
- This program is voluntary.

RANK: ________

ENDNOTES

1. Nonuse values are often defined to include existence, bequest, and intrinsic value.

2. The WTP estimates should consequently not be extrapolated to the population as a whole.

3. Level of knowledge was self-reported.

4. The status-quo or do nothing option was designed to represent current groundwater quality protection efforts.

5. Midpoints of the payment ranges presented in Q-19 were used for this purpose.

6. Mean WTP contingent values were derived from midpoints of payment card ranges, see Q-19.

7. One possibility is that respondents may have felt that an aquifer protection district would not be effective.

8. The avoidance cost approach does not capture losses related to increased fear and anxiety or ecological damages. Moreover, averting expenditures do not provide reliable benefit estimates if they provide other products, such as improved taste.

9. Abdalla's estimate is associated with a 6 month TCE contamination episode in central Pennsylvania.

COMPARING CV RESPONSES WITH VOTING BEHAVIOR: OPEN-SPACE SURVEY AND REFERENDUM IN CORVALLIS, OREGON

Stephen Polasky
Department of Agricultural and Resource Economics
Oregon State University

Olesya Gainutdinova
Joe Kerkvliet
Department of Economics
Oregon State University

May 1996

ABSTRACT

We find that the comparison of the results of the survey and referendum depend on how "undecided" responses are characterized. If "undecided" responses are excluded, there is a statistically significant difference in the percentage of "yes" responses in the survey as compared to the referendum. However, the survey and referendum results can be made to match when 80 to 90% of "undecided" responses are treated as a response of "no." Using the results of the survey, we estimate mean and median willingness to pay for open space of approximately $50 per year (these are preliminary estimates). The estimates of mean and median willingness to pay are not sensitive to the treatment of "undecided" responses.

We thank Helen Berg, Robert Berrens, Charles Vars and participants at the 1996 W-133 Meetings for helpful comments and suggestions. In particular, we thank Robert Mason for help in designing the survey. We also thank Renee Boyd and others at Bardsley & Neidhart, Inc. for their help in implementing the survey.

I. INTRODUCTION

A common justification for using dichotomous choice contingent valuation (CV) is that the decision faced by a CV respondent mimics the decision faced by a voter on whether to vote to provide a certain good at a given (tax) price. A critique of the CV method is that survey responses, being responses to a hypothetical situation, may not match actual responses in a real situation. In order to assess the validity of this criticism, comparisons between CV method results and actual behavior for the same good under the same conditions need to be made. In this paper, we compare responses to a contingent valuation survey with the outcome of a voter referendum on increasing property taxes to buy open space in Corvallis, Oregon, in November 1995.

Critics of the contingent valuation method point out that people may respond quite differently to a hypothetical question than to a real situation (e.g., Diamond and Hausman 1994). There are no actual consequences attached to a hypothetical response, as there are with a real choice. Therefore, an individual may respond to a survey in ways that do not match what they would do in reality. A related issue is that there is less incentive for individuals to acquire information in a hypothetical than in a real setting. As a result, responses to hypothetical questions may be uninformed or poorly thought through.

Because of concerns about the hypothetical nature of survey responses, there is a need to determine the relationship between CV responses and actual behavior. The NOAA panel, charged with assessing the reliability of the CV method, stated:

> External validation of the CV method remains an important issue. A critically important contribution could come from experiments in which state-of-the-art CV studies are employed in context where they can in fact be compared with "real" behavioral willingness to pay for goods that can actually be bought and sold. [Arrow, et al. 1993]

Only a handful of studies have compared hypothetical versus actual responses regarding contributions for specific environmental/public good (Brookshire and Coursey 1987; Kealy, Montgomery, and Dovidio 1990; Duffield and Patterson 1991; Seip and Strand 1992). All four of these studies found evidence that actual contributions were less than stated contributions. (See Cummings and Harrison 1994, pp. 10-11, for a discussion of this issue using the Brookshire and Coursey data).

Other studies have compared CV responses and actual purchase decisions for private goods (e.g., Bishop and Heberlein 1979; Bishop, Heberlein and Kealy 1983; Dickie, Fisher and Gerking 1987; Kealy, Dovidio and Rockel 1988; Neill et al. 1994; Berrens, Adams and Bergland 1994; Cummings, Harrison and Rutström 1995). Many of these studies have found significant response differences in hypothetical versus and real situations. However, the degree to which responses to a hypothetical situation differs from responses to a comparable actual situation and the reasons why this might be so is still a subject of heated debate.

Revealed voting behavior has been suggested as the closest nonexperimental test of the CV method (Hanemann 1994). Deacon and Shapiro (1975) analyzed community voting patterns for two statewide referenda on environmental issues in California. Fischel (1979) surveyed individuals to predict voting patterns on whether to allow siting of a pulp mill. Neither of these two studies attempted to match survey responses with actual voting behavior. In one of the few studies that compare survey responses and voting behavior, Carson et al. (1986) compared responses to a CV survey with referendum results of a bond issue to build sewage treatment plants. They found that the survey results were good predictors of the actual election results, after adjusting for "undecided" respondents.

Comparing CV and voter referenda results has several important advantages over other approaches that attempt to compare responses to hypothetical versus real choices. First, since our study involves a voter referendum, not a simulated market, we avoid questions of whether simulated markets mimic real situations. Second, the referendum offers an explicit implementation rule. If a simple majority of voters vote "yes", the referendum will be enacted. Third, since the payment vehicle is an increase in property taxes, free rider problems associated with voluntary contributions are avoided. Fourth, in comparison with most prior validation studies, the good in question is a public not a private good. Fifth, the referendum concerns an environmental good, the most common subject of the CV method.

In addition to comparing the results of the CV survey with the referendum results, we use the survey result to estimate willingness to pay for open space. How much a person has to pay to purchase open space depends upon the assessed value of property she owns. Assessed value varies widely across individuals in the sample allowing us to estimate a willingness to pay function. The survey also provides evidence on what factors affect willingness to pay that are not available directly from the vote itself.

In the next section of the paper, we describe briefly the open space issue. In section III, we describe the survey and other sources of data. We compare the referendum and survey results in section IV. We then use the survey results alone to estimate a willingness to pay function for open space. The estimation procedure is described in section V and results are given in section VI. Conclusions are given in section VII.

II. THE OPEN SPACE REFERENDUM IN CORVALLIS

Managing growth and preserving open space had become a hot political issue in Corvallis by 1995. For the past several years, Corvallis had grown rapidly fueled by major expansion of a Hewlett-Packard facility and expansion by other area firms. By 1994, according to a survey conducted by the City of Corvallis, half of Corvallis residents felt that Corvallis was "growing too rapidly." There were concerns that specific scenic properties would be developed as well as a general feeling that development would erode the quality of life in the city. Several recent votes on whether to annex land into the city in order to allow development had been defeated by voters. On the other hand, business and real estate interests worried that the open space referendum would drive up the cost of housing in Corvallis and would lead to further measures that would bring a halt to economic growth.

In November 1995, Corvallis voters were asked to decide whether they wished to raise property taxes to fund purchase of open space. The open space measure, if enacted, would allow the city to collect

$950,000 annually for five years for the purpose of purchase of open space. To raise that amount of money, it was estimated that property taxes would increase by 38¢ per $1,000 of assessed value. While priority areas were discussed, specific land parcels to be purchased were not designated as part of the ballot measure. The specific parcels to be purchased would be chosen by the Open Space Advisory Committee, whose members were appointed by the mayor.

There were several other issues on the ballot along with the open space measure: a contest for Benton County sheriff, a proposal to raise property tax to fund an arts center, a proposal to annex some land into the city for development, and a proposal to make minor changes in language to the city charter. The open space measure was the main item of interest on the ballot, generating by far the most vigorous campaign. In the month preceding the election, the open space issue drew approximately 120 letters to the editor of the local newspaper, the *Corvallis Gazette-Times*. This total compares to 30 letters on the sheriff's race and a handful for each of the other issues.

In state and local elections in Oregon, voting by mail rather than the traditional method of voting at a polling place on election day has been allowed. The November 1995 election was the first in Oregon that was conducted exclusively by mail ballots. Ballots were mailed to all registered voters in Corvallis on October 20th. Voters had until 8 pm on November 7th to return their ballots. The length of time between when ballots were mailed and when they had to be returned added an element of complexity to the comparison of survey and voting results, which we discuss further below.

III. DESCRIPTION OF DATA

A. The Open Space Survey

During the evenings of October 16 -19, over 500 telephone interviews of Corvallis residents were conducted by Bardsley & Neidhart, a professional survey research firm based in Portland, OR. The survey was conducted as close as possible to date when ballots were mailed to reduce the possibility that people would change their mind in between the survey and actually voting. The surveys were conducted using random digit dialing of Corvallis phone numbers. Each number was called up to three times to get a response. Respondents were then asked whether they were registered to vote in Corvallis and only those answering "yes" were interviewed. After removing surveys where the respondent did not live in Corvallis, we received 493 completed surveys. A copy of the complete survey is included as appendix A.

After asking a question about how long the respondent had lived in Corvallis and how familiar the respondent was with the open space ballot measure, we asked how the respondent intended to vote on the measure. For half of the surveys, the ballot question exactly as it appeared on the ballot, was read to the respondent:

> "Shall Corvallis acquire open space authorizing a five year tax of $950, 000 annually, outside the constitutional limit?"

We then asked whether the respondent would vote yes or no on the measure. If they were undecided initially, we then asked if they had to vote would they tend to vote yes or no. For the other half of the surveys, we read some background information to the respondent prior to reading the ballot question itself:

> "If the measure passes, it is estimated that property owners will pay 38 cents per thousand dollars of assessed value for 5 years, starting in 1996-97, to purchase open space land. This means that a property owner with a $100,000 house would pay an extra $38 per year in property tax."

The property tax information tells the respondent the price they would face for purchasing open space. Because assessed values vary across respondents, the "tax price" or "bid" level varies across respondents. This variation allows us to estimate a willingness to pay function for open space, which we describe below.

In the survey, we asked respondents whether they owned their own home. If so, we asked them for the assessed value of the home. We also asked whether they owned other property in Corvallis. In addition, we also asked respondents various demographic questions (age, education, income, size of household) as well as how often they personally visited open space or parks and their view of the quality of life in Corvallis.

B. Other Data

We gathered two additional pieces of data on respondents. Using a "reserve phone book" published by US West, we found the name and address for over 50% of the respondents to the survey. Once all of the information of interest to us (whether the respondent voted, respondent's voting precinct, and actual assessed value) are collected, the name, phone number and address information will be deleted from the data base. In some cases, because the phone number was unlisted or various other reasons, a name or address could not be obtained in this manner. For phone numbers that we could track to names, we used voting records for the November election from the Benton County Elections Office to see which respondents had voted in the election. For some phone numbers there were multiple registered voters listed at the address. In this case, we took the total number of registered voters of the same sex as the respondent on the survey and found the proportion of those registered voters who voted. As a result, we derived "fractional votes" for the number of actual voters.

We also obtained actual assessed property value using data from the Benton County Assessor's office matched with the address. We were interested in the degree to which results are sensitive to differences between actual and perceived property values.

IV. COMPARING SURVEY AND REFERENDUM RESULTS

In tables 1 and 2, we compare the percentage of respondents in the survey who said they would vote "yes" on the ballot measure with the actual percentage of "yes" votes in the election. In table 1, we excluded respondents who were "undecided" or "refused" to answer how they intended to vote. In the first row, the results of the referendum are reported. The open space referendum was defeated by a vote of 5,645 (44.77%) "yes" to 6,965 (55.23%) "no." There were only 105 ballots out of 12,715 cast that did not vote either "yes" or "no" on the open space measure. Starting in the second row, we report results from the CV survey. In the second row, we report survey results from the total sample for those who said "yes" or "no" on the ballot question. Out of 493 surveys, there were 127 "undecided" and 10 "refused" to answer the question, leaving 356 surveys with "yes" or "no" responses. Out of those 356 surveys, 53.65% of the respondents said they would vote "yes" on the ballot measure, while 46.35% said "no." The results of the survey and the vote, in terms of the percentage of "yes" responses out of total "yes" and "no" responses, are not very close. A precise statistical test of the hypothesis that the proportions of "yes" responses in the survey and the vote are the same cannot be formulated because we do not know the covariance between survey and vote responses. However, if one is willing to make the reasonable assumption that there is not a negative covariance between survey responses and voting behavior and uses the normal approximation for the binomial distribution, then the hypothesis that the two proportions are the same yields would be rejected at any reasonable level of significance (p-value less than .001).

There are several possible reasons why there are differences between the survey responses and voting behavior. First, not all registered voters cast ballots in the election. If people who were opposed to the measure were more likely to vote, the discrepancy between the survey and voting results could have been caused by the decision of whether or not to vote (i.e., voter selection bias). We checked to see whether voter selection bias was likely to have caused the difference between survey and voting results. In lines 3 and 4 of table 1, we report the results when the sample is restricted to just those whom we could verify as being registered, followed by the subset of those who actually voted in the election. In both cases, the proportion of "yes" responses is almost identical as it is with the original survey. Therefore, it appears that neither our method of tracking registered voters nor voter selection bias is the likely explanation of the difference between the survey and the referendum results.

A second possible explanation for the difference in results between the survey and the referendum is that voters may have changed their mind in between when the survey was taken and when they voted. A typical pattern in referenda is for the proportion of "yes" responses to fall and the proportion of "no" responses to rise as the election comes closer. Magelby (1989, p. 108) reports this pattern in approximately three out of four referenda. Because of the vote by mail election format, there could be a gap of as much as three weeks in between when the survey was taken and when the vote was cast (October 16 was the first day of surveying and November 7 was the last day that a person could have voted). However, the majority of all ballots returned are returned in the first week. Therefore, for the majority of respondents it is likely that the gap between the survey and the vote was approximately one week rather than three weeks. With the existing data, we do not have any good method to test for changes in responses over time. The only evidence relevant to changing opinions over time is the variation in responses over the four days in which the survey was conducted. When we ran a regression that included dummy variables for the date of the survey, there was no discernible time trend over the four day period leading up to the date when the ballots were mailed. Whether there was a longer term trend toward "no" cannot be tested with the available data.

A third possible explanation for the difference in results is that respondents who were "undecided" in the survey voted in a systematic pattern in the election. In the survey, voters who were initially "undecided" were asked if they had to vote would they "tend to vote yes" or "tend to vote no." In the last row of table 1, the results are given for case where we treated those who answered they would "tend to vote yes" as "yes" votes and "tend to vote no" as "no" votes. Again the proportion of "yes" votes in this case is almost identical to the previous cases.

Suppose instead of forcing initially undecided respondents to come down on one side or the other, we treated these respondents as truly undecided at that time. From the survey research literature on political referenda, there is some evidence that most undecided voters close to the election will vote "no." For example, Magelby (1989) states:

> "Because preelection polls are different from actual voting situations in that the actual voting seems to preclude the undecided option, we tested to see whether there was evidence that the "undecided" moves more in the direction of a "no" vote than a "yes" vote. We found that the rate of declining "yes" vote is constant, but the slope of the "no" vote rises, meaning we assume, that at the last minute, the undecided shift overwhelmingly to "no" votes." (p. 107)

In table 2, we report the results from the complete sample, the sample of registered votes, and those who voted, when "undecided" responses are treated as "no" responses. As can be seen from the table, the survey and referendum results are much closer. In order to make the survey results and the referendum results match exactly when the total sample is used, 19.87% of "undecided" responses should vote "yes" with the remainder voting "no." Similarly, for the sample with known registered voters, 13.95% of "undecided" responses should vote "yes" with the remainder voting "no." For the sample with actual voters, this percentage falls to 9.07%. Assigning the vast majority (80-90%) of "undecided" voters as casting "no" votes makes the survey and referendum results match quite well.

A final possibility that we considered is that there may be "yea saying" in the survey responses. Respondents may perceive that saying "yes" is more public spirited or politically correct than saying "no." If so, there may be a fraction of respondents who say "yes" in the survey but vote "no" in the election. The group of respondents who were initially "undecided" may be particularly prone to such behavior since they do not have strong views on the issue to anchor their response. Using a follow-up question to get respondents who were "undecided" to express a tendency to vote "yes" or "no" produced results no closer to the referendum results than did excluding "undecided." However, if we cannot trust the answers to the followup question, can we necessarily trust the answer to the initial question?

In sum, the method that yields the closest prediction between the survey responses and referendum results *in this data* is to assume that 80 to 90% of "undecided" respondents in the survey will vote "no" in the election. The degree to which there is a precise and predictable pattern of responses between surveys and referenda in other cases remains to be demonstrated.

V. ESTIMATION

We use the survey responses to estimate the parameters of a CV willingness to pay function, following Cameron's (1988) censored maximum likelihood procedure. Let WTP_i represent the willingness to pay function for community open space by individual i. This function is defined over a vector of variables, X_i, which includes demographic, attitude and use variables, and an error term, υ_i:

$$WTP_i = X_i'\beta + \upsilon_i.$$

The individual's true WTP_i is an unobservable random variable whose magnitude is inferred through a discrete indicator variable, W_i:

$$W_i = 1 \text{ if } WTP_i \geq \tau_i;\ W_i = 0 \text{ otherwise,} \tag{1}$$

where τ_i is the bid level or tax price. In either case, τ_i is the variable to which the individual reacts, accepting or rejecting if her true willingness to pay is above or below this bid threshold.

Using information on the probability distribution of W_i across levels of τ_i, the parameter vector, β, of the WTP function can be estimated. Following Cameron (1988), assume υ_i is a logistic random variable with mean zero and standard deviation $b = \pi/3$ (and dispersion parameter $k= b\sqrt{3}/\pi$). The probability of answering "yes" when answering the valuation question is:

$$P = \text{Prob}(W_i = 1) = \text{Prob}(WTP_i \geq \tau_i)$$
$$= \text{Prob}(\nu_i \geq \tau_i - X_i'\beta)$$
$$= \text{Prob}((\nu_i /k) \geq (\tau_i - X_i'\beta)/k) \ . \quad (2)$$

Given the assumptions about the random variable υ_i, P takes the form:

$$P = [1 + e^{(\tau_i - X_i'\beta)/k}]^{-1} \ ; \quad (3)$$

and for a "no" response:

$$(1-P) = [e^{(\tau_i - X_i'\beta)/k}] \cdot [1 + e^{(\tau_i - X_i'\beta)/k}]^{-1} \ . \quad (4)$$

Combining equations (4) and (5), the likelihood function is:

$$L = \Pi_{W_i=1}[(P)] \ \Pi_{W_i=0}[(1-P)] \ , \quad (5)$$

which by substituting (3) and (4) can be expressed as:

$$L = \Pi_{W_i=1}[(e^{(\tau_i - X_i'\beta)/k})^{-1}] \ \Pi_{W_i=0}[\frac{e^{(\tau_i - X_i'\beta)/k}}{1 + e^{(\tau_i - X_i'\beta)/k}}] \ . \quad (6)$$

Nonlinear optimization methods can be used to find maximum likelihood estimates of 1/k and β and their standard errors, and, given X, estimates of each respondent's WTP.

In this model, we assume that the eligible voter responds to this question based on whether or not their willingness to pay is greater than or less then the private tax consequences of passage of the referendum. Hence, τ_i is taken to be the homeowner's tax price, defined as the product of the home's assessed value and the referendum's proposed mill levy increase. In estimating this model, we wish to avoid problems with probable differences in perceived tax prices for homeowners, renters, and business owners (Oates, 1994). Our sample will be limited to respondents who own their own homes. Ownership of other real estate is controlled for either by using an indicator variable or by further restricting the sample to home owners who do not own other real estate in the community.

VI. WILLINGNESS TO PAY ESTIMATES

Using the methods outlined in the previous section, we estimate a logistic regression equation where the dependent variable is equal to one for a "yes" response and equal to zero for a "no" response. For all the results reported in this section, we restrict the sample population to property owners. In the initial set of results reported in table 3, we further excluded respondents who were "undecided." In table 4, we coded "undecided" responses as "no." It should be stated that the results in this section are preliminary. We plan to carry out a more detailed and careful analysis of the data in the near future.

We report results using all homeowners that stated an opinion on the open space question as well as for those that voted. We used two different measures for assessed value of the respondent's residence: i) actual assessed value of the property, ii) the assessed value response from the survey. In the survey, only categories of assessed values where given. We coded responses in a category to be equal to the midpoint of the range of the category. The other independent variables are defined as follows:

- Other property: a dummy variable that equals one is the respondent owned other property in Corvallis in addition to their place of residence;
- Parks/open space use: a dummy variable that equals one if the respondent visited a park or open space in the Corvallis on average of at least once a month;
- Quality of life: how the respondent rated quality of life in Corvallis on a scale of 1 to 10 with 10 being "very satisfied";
- Version A: a dummy variable that equaled one if the respondent was not read any additional information besides the ballot measure;
- Income: household income (by category);
- Education: highest level of education obtained by the respondent (by category);
- Number of adults: number of persons 18 years or older living in the household;

- Number of children: number of person less than 18 years old living in the household;
- Male: dummy variable that equaled one if the respondent was male.

In future work, we plan to restrict the sample to homeowners that do not own other property as an alternative to including homeowners with other property but controlling for the effect of other property with a dummy variable.

The results of the regression runs are reported in tables 3 and 4. Generally, results with "undecided" coded as "no" (table 4) yield less precise estimates than when undecided are excluded (table 3). Further, since the sample of actual voters is a subset of the complete sample, it is harder to find significant coefficients with the sample of actual voters. Of greatest interest to us is the coefficient on the variable of assessed house value. The coefficient has the expected negative sign in each regression except in column 2 of table 4. This case uses only respondents who we could track as having voted, codes "undecided" respondents as voting "no," and used actual rather than perceived assessed value. Using perceived as opposed to actual assessed value generates a larger negative coefficient on assessed value. In other words, increases in perceived assessed value as opposed to actual assessed value are more likely to yield an increase in the percentage of "no" responses. In column 1 of table 5, we report the change in probability of responding "yes" with a change in assessed value for the various regression results presented in tables 3 and 4. The decline in probability of responding "yes" with an increase in house value of $100,000 ranges from a high of almost 22% to an increase of 1.8%. As stated previously, an increase in assessed value of $100,000 translates to an approximate $38 increase in taxes per year.

Using the regression results, we can obtain estimates of annual mean and median willingness to pay for open space, which are reported in columns 2 and 3 of table 5. The results over all cases turned out to be quite stable. For annual mean willingness to pay, the estimates vary between $51.99 and $54.54. For annual median willingness to pay the estimates vary between $43.51 and $54.92. These estimates of annual willingness to pay seem quite reasonable. The estimated mean and median willingness to pay were not sensitive to how "undecided" responses were treated. This contrasts with the conclusion from section IV where the ability to predict the outcome of the referendum from survey data was quite sensitive to the how "undecided" responses were treated.

Several other results are worth noting. Whether or not we gave respondents additional information about how to interpret how much the open space referendum would likely cost them did not seem to influence the likelihood of responding "yes" to the open space question. In the complete sample, reading additional information increased the probability of responding "yes." However, in the sample restricted to those who voted, reading additional information reduced the probability of responding "yes." Neither result was statistically significant at even a 10% significance level. Another variable that seemed to have little influence on the likelihood of responding "yes" was income. The coefficient on income had a negative sign in three out of eight cases and a positive sign in five out of eight cases. The coefficient on income was never statistically significant. The two variables that were consistently positively related to a "yes" response were education and the respondent's use of parks and open space.

VII. CONCLUSIONS

We find that the comparison of the results of the survey and referendum depends on how "undecided" responses are characterized. If "undecided" responses are excluded, there is a statistically significant difference in the percentage of "yes" responses in the survey as compared to the referendum. However, the survey and referendum results can be made to match when 80 to 90% of "undecided" responses are treated as a response of "no." Using the results of the survey, we estimated mean and median willingness to pay for open space of approximately $50 per year. The estimates of mean and median willingness to pay were not sensitive to the treatment of "undecided" responses. There should be caution in the interpretation of the willingness to pay results. Most referenda, just like most surveys, do not capture a pure expressions of willingness to pay for a certain good. Rather, a vote or a survey about whether to increase taxes to pay for an environmental good may embody opinions about government and taxes as well as the value of the environmental good in question.

TABLE 1
Comparing Survey and Referendum Results without Undecided and Refusals

	Observations	Yes	No	Percent Yes
Referendum Vote	12610	5645	6965	44.77
Total Sample - Yes/No	356	191	165	53.65
Registered Voters	202	108	94	53.47
Actually Voted	162.75	87.25	70.5	53.61
Total Sample with Tend Yes and Tend No	423	229	194	54.14

TABLE 2
Comparing Survey and Referendum Results Interpreting Undecided and Refusals as No

	Observations	Yes	No	Undecided	Percent Yes
Total Sample	483	191	165	127	39.54
Registered Voters	259	108	94	57	41.7
Actually Voted	203.05	87.25	75.5	40.3	42.97

TABLE 3
Logistic Regression Results: Excluding Undecided

Variable	1. All Homeowners Actual Assessed Value	2. Voters Actual Assessed Value	3. All Homeowners Self-Reported Assessed Value	4. Voters Self-Reported Assessed Value
Constant	3.9598 (2.4001)	4.2011 (3.0526)	4.4692 (2.4253)	5.7940 (3.1151)
House Value ($000)	-0.004704 (0.004717)	-0.002552 (0.006949)	-0.007207 (0.005201)	-0.01482 (0.008128)
Other Property	-1.5376 (0.9650)	-1.1537 (1.2655)	-1.6713 (0.9602)	-1.5976 (1.2530)
Parks/Open Space Use	1.7791 (0.4372)	1.9116 (0.5687)	1.6751 (0.4413)	1.6389 (0.5697)
Quality of Life	0.1840 (0.1271)	0.1160 (0.1783)	0.1876 (0.1271)	0.2172 (0.1874)
Version A	-0.3741 (0.3823)	0.1122 (0.5020)	-0.3762 (0.3855)	-0.01605 (0.5161)
Income	0.05209 (0.1602)	-0.05962 (0.2149)	0.07526 (0.1588)	0.06680 (0.2206)
Education	0.6774 (0.1741)	0.7370 (0.2250)	0.6838 (0.1740)	0.8338 (0.2409)
Number of Adults	-0.6199 (0.3510)	-0.8759 (0.4723)	-0.6164 (0.3519)	-0.8684 (0.4737)
Number of Children	0.2819 (0.1856)	0.4642 (0.2752)	0.3461 (0.1968)	0.5827 (0.2836)
Male	-0.5450 (0.4079)	-0.5328 (0.5543)	-0.5474 (0.4092)	-0.6808 (0.5727)
Log Likelihood Function	-83.8007	-50.3993	-82.4107	-48.6765
Percent Correct Predictions	72.56	78.5	73.91	78.5
Number of Observations	164	107	161	107

TABLE 4
Logistic Regression Results: Including "Undecided" as "No"

Variable	1. All Homeowners Actual Assessed Value	2. Voters Actual Assessed Value	3. All Homeowners Self-Reported Assessed Value	4. Voters Self-Reported Assessed Value
Constant	1.6605 (1.9596)	1.3415 (2.5017)	2.4550 (2.0213)	3.1644 (2.6327)
House Value ($000)	-0.0009919 (0.003805)	0.0009770 (0.005171)	-0.005571 (0.004466)	-0.01183 (0.006619)
Other Property	-0.6490 (0.7050)	-0.1373 (0.9069)	-0.8672 (0.7123)	-0.7382 (0.9499)
Parks/Open Space Use	1.3487 (0.3932)	1.5056 (0.5127)	1.1963 (0.3965)	1.1974 (0.5153)
Quality of Life	0.07800 (0.1162)	0.0541 (0.1695)	0.0976 (0.1163)	0.1649 (0.1758)
Version A	-0.2520 (0.3316)	0.1122 (0.5020)	-0.2740 (0.3345)	-0.005274 (0.4418)
Income	-0.02822 (0.1347)	0.1395 (0.4335)	0.02747 (0.1332)	-0.02222 (0.1853)
Education	0.4967 (0.1484)	0.6102 (0.2008)	0.5201 (0.1510)	0.6975 (0.2117)
Number of Adults	-0.2895 (0.2940)	-0.4889 (0.3944)	-0.2929 (0.2929)	-0.4559 (0.3906)
Number of Children	0.2019 (0.1568)	0.3953 (0.2297)	0.2606 (0.1647)	0.4946 (0.2345)
Male	-0.2647 (0.3444)	-0.0608 (0.4558)	-0.2899 (0.3464)	-0.2212 (0.4718)
Log Likelihood Function	-109.1	-66.3788	-107.036	-64.7035
Percent Correct Predictions	68.45	70.25	68.31	75.21
Number of Observations	187	121	183	121

Table 5
Change in Probability of a "Yes" Response with a Change in Assessed Value, Mean and Median Willingness to Pay

Case	1. Marginal Value	2. Mean Willingness to Pay	3. Median Willingness to pay
"Undecided" excluded, complete sample, actual assessed value	0.0007959	$52.69	$51.30
"Undecided" excluded, voters only, actual assessed value	0.0003901	54.54	52.08
"Undecided" excluded, complete sample, perceived assessed value	0.001224	53.41	54.92
"Undecided" excluded, voters only, perceived assessed value	0.002184	53.04	43.51
"Undecided" as "no", complete sample, actual assessed value	0.0001976	51.99	46.7
"Undecided" as "no", voters only, actual assessed value	-0.0001805	54.03	52.23
"Undecided" as "no", complete sample, perceived assessed value	0.001114	53.32	54.91
"Undecided" as "no", voters only, perceived assessed value	0.002123	53.23	43.51

REFERENCES

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner and H. Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register* 58: 4601-4614.

Berrens, R., R. Adams and O. Bergland. 1994. Information Transfer and Revenue Capture Using Contingent Valuation: Pheasant Hunting in Oregon Revisited. Working Paper, Department of Agricultural and Resource Economics, Oregon State University.

Bishop, R.C. and T.A. Heberlein. 1979. Measuring Values of Extra-Market Goods: Are Indirect Methods Biased? *American Journal of Agricultural Economics* 61: 926-930.

Bishop, R.C., T.A. Heberlein and M.J. Kealy. 1983. Contingent Valuation of Environmental Assets: Comparison with a Simulated Market. *Natural Resources Journal* 23: 619-634.

Brookshire, D. and D. Coursey. 1987. Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures. *American Economic Review* 77: 554-556.

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

Carson, R.T., W.M. Hanemann, and R.C. Mitchell. 1986. The Use of Simulated Political Markets to Value Public Goods. Economics Department, University of California at San Diego.

Cummings, R.G. and G.W. Harrison. 1994. Was the Ohio Court Well Informed in its Assessments of the Accuracy of the Contingent Valuation Method? *Natural Resources Journal* 34: 1-36.

Cummings, R.G., G.W. Harrison and E.E. Rutström. 1995. Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible? *American Economic Review* 85: 260-266.

Deacon, R. and P. Shapiro. 1975. Private Preference for Collective Goods Revealed through Voting on Referenda. *American Economic Review* 65: 943-955.

Diamond, P.A. and J.A. Hausman. 1994. Contingent Valuation: Is Some Number Better than No Number. *Journal of Economic Perspectives* 8: 45-64.

Dickie, M., A. Fisher, and S. Gerking. 1987. Market Transactions and Hypothetical Demand Data: A Comparative Study. *Journal of the American Statistical Association* 82: 69-75.

Duffield, J.W. and D.A. Patterson. 1992. Field Testing Existence Values: Comparison of Hypothetical and Cash Transaction Values. Fifth Interim Report of the USDA Western Regional Research Project W-133.

Fischel, W.A. 1979. Determinants of Voting on Environmental Quality: A Study of a New Hampshire Pulp Mill Referendum. *Journal of Environmental Economics and Management* 6: 107-118.

Hanemann, W.M. 1994. Valuing the Environment through Contingent Valuation. *Journal of Economic Perspectives* 8: 19-43.

Kealy, M.J., J.F. Dovidio and M.L. Rockel. 1988. Accuracy in Valuation is a Matter of Degree. *Land Economics* 64: 158-171.

Kealy, M.J., M. Montgomery and J.F. Dovidio. 1990. Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter? *Journal of Environmental Economics and Management* 19: 244-263.

Magelby, D. 1989. Opinion Formation and Opinion Change in Ballot Proposition Campaigns. In *Manipulating Public Opinion*, M. Margolis and G Mauser (eds.). Pacific Grove, CA: Brooks/Cole Publishers.

Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison and T. McGuckin. 1994. Hypothetical Surveys and Real Economic Commitments. *Land Economics* 70: 145-154.

Oates, W. 1994. Estimating the Demand for Public Goods: The Collective Choice and Contingent Valuation Approaches. Department of Economics, University of Maryland.

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: 91-106.

Appendix A

VERSION B CORRVALLIS OPEN SPACE SURVEY 10 OCT 95

1. Are you registered to vote at your current address in Corvallis, or not?

YES 1
NO . . . . . 2 **(TERMINATE)**

2. How many years have you, yourself, lived in Corvallis?

YEARS __________
DK/NA 99

3. Some people don't pay much attention to local elections while others are interested. Have you heard about an election next month in which people will vote on using public funds to purchase open space?

YES, HEARD . . . 1
NO 2
DK/NA 3

4. As you may know, we will be voting on a ballot measure in November that I would like your opinion on. If the measure passes, it is estimated that property owners will pay 38 cents per thousand dollars of assessed value for 5 years, starting in 1996-97, to purchase open space land. This means that a property owner with a $100,000 dollar house would pay an extra 38 dollars per year in property tax. The ballot title reads:

"Shall Corvallis acquire open space authorizing a five-year tax of $950,000 annually, outside the constitutional limit?" Would you vote YES or NO on this measure if the election were held today?"

VOTE YES **(SKIP TO Q 4b)** . . 1
VOTE NO **(SKIP TO Q 4b)** . . 2
UNDECIDED/DK **(GO TO Q 4a)** . 3
REFUSED **(SKIP TO Q 5)** . . . 4

4a. If you had to vote would you tend to vote yès or no on the open space measure?

STILL UNDECIDED/DK/NA **(SKIP TO Q 5)** . 1
TEND TO VOTE YES **(GO TO Q 4b)** . . . 2
TEND TO VOTE NO **(GO TO Q 4b)** . . . 3

4b. Would you please tell me why you would vote (yes)(no) on the open space measure? **(PROBE!)**

What else?

5. How often do you visit any park or open space in or near Corvallis -- daily, weekly, monthly, several times a year, once a year, or never?

DAILY 1
WEEKLY . . . 2
MONTHLY . . . 3
SEVERAL . . . 4
ONCE 5
NEVER 6

6. On a scale of 0 to 10, in which a "0" is **"very dissatisfied"** and "10" is **"very satisfied"** how would you rate the quality of life in Corvallis?

RATING __________
DK/NA 99

Finally, I would like to ask you a few questions about yourself so we can check our sample. All information is, of course, strictly confidential.

7. First, I'd like to read you some broad age groups. Please stop me when I come to the group in which you belong.

18-24 . . . 1
25-34 . . . 2
35-44 . . . 3
45-54 . . . 4
55-64 . . . 5
65+ 6
Refused . . 7

8. What was the last grade you completed in school and got credit for?

GRAD/PROF 1
COLLEGE-COMPLETE . . 2
COLLEGE-PARTIAL . . . 3
COMM. COLL, TRADE . . 4
COMPLETE HIGH SCHOOL 5
LESS THAN HS 6
OTHER (_____________) 7
REFUSED 8

9. And may I ask do you own or rent the dwelling unit in which you now live?

REFUSED **(SKIP TO Q 10)** 1
PROVIDED **(SKIP TO Q 10)** 2
RENT **(SKIP TO Q 10)** . . 3
OWN **(GO TO Q 9a)** . . . 4

9a. I would like to read some broad home assessment groups. When I come to the one that most closely represents the assessment for your house, will you please stop me? **(JUST YOUR BEST ESTIMATE, PLEASE)**

LESS THAN $70,000 0
$70,000 - $99,999 1
$100,000 - $129,999 2
$130,000 - $159,999 3
$160,000 - $189,999 4
$190,000 - $219,999 5
$220,000 - $249,999 6
$250,000 - $299,000 7
$300,000 OR GREATER 8
UNCERTAIN 9
REFUSED 10

10. Do you own other real estate in Corvallis?

YES 1
NO 2
REFUSED . . 3

11. Including yourself, how many adults 18 years or older are living in your household at the present time?

NUMBER ______
REF . . 99

12. And, how many children less than 18 years old, if any, are living in your household at the present time?

NUMBER ______
REF . . 99

13. I would like to read you some broad income groups. When I come to the one that best represents your total household income before taxes, will you please stop me? **(JUST YOUR BEST ESTIMATE)**

UNDER $15,000	1
$15,000 - $24,999 . . .	2
$25,000 - $34,999 . . .	3
$35,000 - $49,999 . . .	4
$50,000 - $74,999 . . .	5
$75,000 - $99,999 . . .	6
$100,000 AND OVER . . .	7
REFUSED	8

14. (INT: RECORD GENDER NOW)

MALE . .	1
FEMALE .	2

VERIFY PHONE NUMBER NOW: ____________________

Refused 999-9999

(THANK YOU VERY MUCH FOR YOUR TIME!)

A COMPARISON OF CVM AND POINT ALLOCATION APPROACHES TO ESTIMATING NON-USE VALUES FOR WILDERNESS AREAS*

John E. Keith
Christopher Fawson
Van Johnson
Department of Economics
Utah State University
Logan, Utah 84322-3530

ABSTRACT

Both point allocation and total and use values for wilderness designation in Utah were compared to examine use and non use values. Results indicated that use, option value, existence value and bequest value have statistically almost identical allocations across groups favoring wilderness in general, the Bureau of Land Management proposal for wilderness, and the Utah Wilderness Coalition proposal. However, when the difference between total and use values calculated from willingness to pay measures is compared to the point allocations, there does not appear to be a consistent relationship between use and non-use values; that is, option value cannot clearly be defined as either use or non-use.

---

*This study was in part supported by funds from the W-133 Regional Project and the Utah Agricultural Experiment Station.

INTRODUCTION

The determination of the non-use values of public recreational goods has been the subject of several studies and articles. Greenly, Walsh and Young (1981) identified four parts of value (use, option, existence, and bequest) and estimated them using contingent valuation techniques wherein the respondents were asked to allocate portions of their total willingness to pay(WTP) for hypothetical wilderness areas to actual use, possible future use, satisfaction of knowing the wilderness areas existed, and satisfaction of knowing the areas would be protected for future generations. Following this article, and others concerning option value and existence value (McConnell; Walsh, Loomis and Gillman; and Smith), there seems to be a general consensus that "...'option value' does not represent a distinct component of value, and that...total value...{is composed of}...use values (e.g., consumptive, nonconsumptive, and indirect), with what remains being termed existence value or nonuse value." (Larsen)

THE STUDY

As a part of a larger examination of wilderness designation in Utah, a contingent valuation (CV) study to determine Utahn's willingness to pay for wilderness designation or non-designation was completed. While several proposals for wilderness have been made in Utah, only two have well-documented specific proposals:[1] the Utah Wilderness Coalition (UWC) recommendation which was published as <u>Wilderness at the Edge</u>, and the U. S. Bureau of Land Management (BLM) recommendation, which was reported in their related Final Environmental Impact Statement. The former proposal comprises approximately 5.7 million acres of BLM land; the latter, about 1.9 million acres. Most of this acreage is, quite obviously, in sparsely populated rural Utah, which has depended heavily on a traditional extractive resource economic base.

The contingent valuation portion of the study focused on these two proposals for wilderness designation. A sample of 2,135 Utah households was drawn by Survey Research, Inc. of Arlington, Virginia. Given the distribution of Utah population (over 80 percent of the population resides in the urbanized Wasatch Front), a second sample (600 households) of only the rural population was requested, in the anticipation that some rural counties would not be adequately represented in the general population sample. The samples used in the study included the original general population respondents, the urban respondents (those households residing in counties along the Wasatch Front, plus Cache County) and rural respondents (all other counties).

A computer-based contingent valuation questionnaire was developed for the study. Prior to its implementation, a packet of information was sent to each household in the sample. That packet included a map detailing the existing wilderness areas in Utah and the two proposals for designation along with a brief explanation of the regulations which have been implemented to constrain the use of recently designated wilderness areas.[2] These regulations include clauses which specify that no reduction in existing traditional uses will be made unless the "wilderness quality" is threatened, and that the traditional means of extraction will be allowed (for example, trucking and mechanized maintenance for grazing). However, further development of grazing, minerals, or other traditional extractive uses, and the use of mechanized recreational equipment, are prohibited.

In addition to the information regarding the wilderness proposals, a letter explaining the contingent valuation study and information concerning the survey itself was included. This letter indicated that willingness to pay for designation or non-designation would be collected, along with attitudinal and socioeconomic data.

Upon telephone contact with a household, the interviewer asked for the person 18 or older in the household who had most recently had a birthday. This approach was used to help insure a random sample. If the packet had not been received read or retained by the respondent, a new packet was sent and the individual re-contacted at a later date.

The respondent was first asked about his/her past history of visitation to wilderness areas in Utah and participation in various kinds of outdoor recreation activity. He or she was next asked to rank his/her feelings about wilderness in general using a Likert scale of 1 to 10, where 1 signified strong opposition; 5, neutrality or no opinion; and 10, strong support. If the respondent gave a ranking of 5 or above, he or she was classed as "supporting" wilderness in general; a ranking of 4 or less resulted in being classified as "opposing" wilderness in general. Supporters of "wilderness in general" were then asked to allocate 10 points to each of four categories of reasons for his/her support, as follows:

> `Now, suppose you have 10 points to allocate among reasons why you favor wilderness areas in general. You may allocate all 10 points to one reason, or divide them up according to your feelings about the relative importance of each reason. I will read the reasons, and then ask you to give me your allocation. Remember that the total must add up to 10.
>
> A. I or members of my family will use these wilderness areas and want them for my continued use.
>
> B. There is a chance that I or members of my family will use these areas, and I would like to have them available if and when I decide to use them.
>
> C. I would like to have these areas available for others to use even if I or members of my family never use them.
>
> D. I would like to have these areas available for future generations to use, even if I or members of my family never use them.

In the event that an individual did not allocate exactly 10 points, he or she was reminded about the limit of points, and the questions were asked again. This allocation question was asked only once, after the initial "support of wilderness in general" question. It was assumed that individual respondents would make those same allocations for any wilderness proposal.

After the allocation questions were completed, a series of CV questions regarding use of the existing wilderness areas was then asked.[3] Next, the same ranking criteria were applied to the BLM proposal, followed by a set of CV questions which addressed both total value (establishment of the wilderness areas proposed by the BLM) and use value (annual WTP for use of the area). Lastly, those criteria were applied to the UWC proposal and the final set of CV questions were asked. After the CV questions were completed, information on the socioeconomic characteristics of the respondent and his or her spouse (if any) were collected, including race, age, education, employment, marital status, and income. The telephone interviews lasted, on average, about 20 minutes.

A closed ended dichotomous choice approach was used for the CV questions. This approach is generally (although not entirely) accepted as the standard approach for minimizing various kinds of bias in CV studies.[4] The "bid values" were chosen based on earlier work in Utah by Pope and Jones, on other wilderness studies as reported by Walsh, et al., and on a pre-test using a non-random sample of individuals on the Utah State University campus. Those values ranged between $10 and $2,000, and were selected at random by the computer program used for the survey. The CV questions for the proposed designations (BLM and UWC) involved two "steps." First, the individual was asked a referendum question about his or her willingness to vote for the designation of the BLM (or the UWC) proposed areas given that designation would result in a specified annual income loss (in perpetuity). The bid represented a measure of total value. Next, the respondent was asked about his or her willingness to pay for an annual permit to use those areas as wilderness, using the same format as used for the use of existing wilderness areas. Those responding that they would vote against designation and/or would not pay any fee were asked to specify their reason for not being willing to pay. These reasons were classified as economically-based ("not worth it to me," or "I can't afford it") or protests ("I shouldn't have to pay for wilderness," "I object to the payment (or the question)," etc.).

The willingness to pay measure (compensating surplus) was estimated for each question and each group of non-protest respondents (general wilderness, BLM, and UWC). Two alternative approaches were used. The first followed Hanemann (1984, 1989), using a logit estimator and the linear form of the indirect utility model. This approach admits negative responses (that is, part of the density function may be found in the negative quadrant, indicating a positive probability of a negative willingness to pay). This "negative response" is frequently found in CV studies; ours was no exception. Given the possibility of negative responses in a case in which there is clearly no expectation of negative responses (such as ours in which opponents were eliminated from the estimation), one can choose to truncate the distribution, which biases the estimated WTP, or to use a log-linear form which is consistent with only positive bids (note that the log linear form excludes 0 bids), such as the estimations by Bishop and Heberlein. Johansson, et al., in their comment on Hanemann's 1984 article, briefly suggest that this form, and others which are local approximations to utility functions, might be used. We estimated WTP using both the truncated linear and the log-linear models.

STUDY RESULTS

Point allocation results are listed in Tables 1, 2 and 3. The standard deviations for each "reason" are relatively large, although the means are remarkably consistent for all three cases. The distribution of attitudinal values followed the distribution found in other attitudinal studies (Table 4): relatively large numbers of "5" scores, scores between 5 and 9 centered on the median scores (7 and 8), and a relatively large number of "10" values. The correlations between the attitudinal measure (5 to 10 for supporters of wilderness) and the point allocation for non-protest respondents for both the BLM and UWC proposals are found in Table 5. In both cases, the higher the attitudinal score, the higher were the scores in both the use and bequest values, but the lower were the scores in option and existence values. Further, use values appear to have a negative correlation with all other categories, and a relatively strong negative correlation (-.50) with bequest values. Simple OLS analysis for both proposals suggests a significant ($t \geq 2.0$; .025 level) negative coefficient between the attitude measure and option value, and a somewhat less significant ($t \geq 1.75$; .05 level) positive coefficient between the attitude measure and bequest value. The coefficients for existence (negative) and use (positive) were not significant except for a slightly significant ($t = 1.55$; .1 level) positive relationship for BLM use allocations.

Table 1. Point Allocation for Wilderness in General

	USE	OPTION	EXISTENCE	BEQUEST
All Wilderness Supporters				
Total	1,509	992	1,182	2,177
Per Cent	25.75	16.93	20.17	37.15
Std Dev	22.88	16.74	17.09	25.08
Sum	42.68		57.32	
Non-Protest Supporters				
Total	908	590	719	1,383
Per Cent	25.22	16.39	19.97	38.42
Std Dev	22.52	15.57	16.69	25.01
Sum	41.61		58.39	

Table 2. Point Allocation for BLM Wilderness Proposal

	USE	OPTION	EXISTENCE	BEQUEST
All Supporters				
Total	1,157	753	916	1,734
Per Cent	25.37	16.51	20.09	38.03
Std Dev	22.36	16.38	17.13	24.72
Sum	41.88		58.12	
Non-Protest Supporters				
Total	883	503	633	1,231
Per Cent	27.17	15.48	19.48	37.88
Std Dev	23.55	15.32	16.66	24.86
Sum	42.65		57.35	

Table 3. Point Allocation for UWC Wilderness Proposal

	USE	OPTION	EXISTENCE	BEQUEST
All Supporters				
Total	959	593	755	1,433
Per Cent	25.64	15.86	20.19	38.32
Std Dev	22.53	16.06	17.45	25.49
Sum	41.50		58.50	
Non-Protest Supporters				
Total	680	413	532	1,055
Per Cent	25.37	15.41	19.85	39.37
Std Dev	22.55	16.73	18.53	26.32
Sum	40.78		59.22	

Table 4. Distribution of Attitudinal Score

	BLM		UWC	
	Number	Pct	Number	Pct
5	81	23.9	85	29.9
6	33	9.7	35	12.3
7	53	15.6	40	14.1
8	68	20.1	49	17.3
9	23	6.8	8	2.8
10	81	23.9	67	23.6

Table 5. Correlation Coefficients

BLM PROPOSAL

	Attitude	Use	Option	Existence	Bequest
Attitude	1				
Use	0.08887	1			
Option	-0.25337	-0.23342	1		
Existence	-0.09750	-0.38025	0.03534	1	
Bequest	0.13728	-0.54855	-0.41876	-0.33180	1

UWC PROPOSAL

Attitude	1				
Use	0.01212	1			
Option	-0.14499	-0.20079	1		
Existence	-0.02942	-0.31449	-0.03952	1	
Bequest	0.10253	-0.50712	-0.43619	-0.40964	1

The results from the linear logit estimations for non-protest bids are found in Table 6 along with the number and percentage of each group. These results appear consistent with expected results. It should be noted that the number of non-protest bids changed the number of observations used in the estimates. In addition, many of the intercepts and coefficients for the income variable are not statistically significant at the 10 or even the 20 percent level. All of the coefficients for the "bid" are statistically significant at least at the 10 percent level. For the non-protest segment of the respondents, the percentage of those supporting wilderness designation declines between the smaller BLM and larger UWC proposals. Table 7 indicates the WTP values which are obtained from the approach discussed above (Hanemann, 1984, 1989) for the simple linear random utility model which does not admit negative values ($y[1 - e^{-\alpha/\beta}]$), and bootstrapped confidence intervals (Cooper, 1994).

Table 6. Logistic results (t statistics in parentheses)

Wild	Establishment			Use				
Prop	Int	Bid	Inc	Int	Bid	Inc	N	Pct
WLD Suprt	N/A	N/A	N/A	.01898 (.07)	-.00293 (-5.58)	.02010 (.38)	365	90.0
BLM Suprt	.3405 (1.17)	-.00119 (-5.29)	-.00286 (-.05)	1.7906 (4.74)	-.00898 (-7.21)	-.04682 (-.66)	316	72.3
UWC Suprt	.3228 (1.06)	-.00209 (-5.54)	.072 (1.11)	1.5724 (4.53)	-.00564 (-6.72)	-.0961 (-1.41)	284	63.6

Table 7. Calculated willingness to pay by proposal - linear model, negative WTP not admitted (Confidence intervals in parentheses)

Proposal	Establishment	Use
WLD Support	N/A	$ 255 ($214-292)
BLM Support	$ 729 ($568-$842)	$ 198 ($173-$216)
UWC Support	$ 498 ($392-$556)	$ 254 ($202-$281)

The percentage of use value of total is 27.16 percent for the BLM case and 51.00 percent for the UWC case. The former value is very close to the use category of the point allocation; however, the latter value is closer to the use category plus the option category. In general, there does not seem to be, at least for the general population, a clear indication of the definition of use (that is, with or without option value). When urban and rural respondents were examined separately, the results stayed essentially the same.

The log-linear (in price) logit estimations are presented in Table 8. Because the log-linear indirect utility function has no closed form solution for willingness to pay, and because the value of the coefficients on the bid variable can result in infinite willingness to pay when the entire distribution is considered, Bishop and Heberlein (1979) and Johansson, et al.(1989), have suggested an estimate truncated at the maximum bid amount. This avoids the effects of the "fat" tails of the logistic estimation. This is a particular problem for cases of "yea-saying" with high values of bids. Of course, this truncation underestimates the mean willingness to pay derived from the linear approach. The calculated willingness to pay for the log-linear logit models is found in Table 9. When an upper limit to the integration yielding WTP was $3,000, the results exceeded the maximum bid by substantial amounts. Therefore, we used an upper limit of the maximum bid, $2,000, as suggested above. Confidence intervals were again obtained using bootstrapping methods.

A comparison of the linear and the log-linear results suggests some substantial differences, and some inconsistencies. Note that log-linear results are generally larger than the linear models for the supporters of wilderness for both establishment and use values. Furthermore, for both the BLM and UWC cases, the use values exceed establishment (total) values. This result is inconsistent, since total value should include use <u>and</u> existence values. Note that the linear results exhibit the expected relationship between establishment and use values. These results seem to suggest that at least the log-linear model should not be used for dichotomous choice questions, although further examination of our data as well as other tests are needed to confirm our results.

SUMMARY AND CONCLUSIONS

The study appears to substantiate other researchers' results that the division of use and non-use values is problematic. Our results may stem from the fact that respondents were not asked to allocate points for each of the alternative proposals, as well as for wilderness in general. However, one would

Table 8 Logistic log-linear results for the general population sample (t statistics in parentheses)

Wild	Establishment			Use				
Prop	Int	Bid	Inc	Int	Bid	Inc	N	Pct
WLD Suprt	N/A	N/A	N/A	3.1022 (5.52)	-.77537 (-7.31)	.00806 (.14)	365	90.0
BLM Suprt	2.2104 (4.68)	-.48334 (-6.35)	.00447 (.07)	7.2382 (8.66)	-1.4536 (-9.39)	-.08194 (-1.10)	316	72.3
UWC Suprt	2.9690 (5.58)	-.70638 (-7.46)	.09553 (1.43)	6.2623 (8.21)	-.1.214 (-8.70)	-.1299 (-1.80)	284	61.1

Table 9. Calculated willingness to pay by proposal - log-linear model (Confidence intervals in parentheses)

Proposal	Establishment	Use
WLD Support	N/A	$1,691 ($1,624-$1,748)
BLM Support	$1,348 ($1,292- $1,484)	$1,762 ($1,706- $1,786)
UWC Support	$1,510 ($1,415 $1,595	$1,718 ($1,631 $1,760)

suspect that the larger the proposal, the more likely that existence or bequest values would increase. In any case, the inclusion of option value as a part of use value does not seem warranted by the study's result, if the point allocation process accurately reflects respondents' reasons for supporting wilderness. Secondly, our results from the log-linear model indicate that choosing a local approximation to a utility function in order to avoid the truncation inherent in Hanemann's linear model may result in inconsistent values. Non-parametric estimations of the distributions of responses may furnish a more robust approach to dichotomous choice contingent valuation estimations.

LITERATURE CITED

Bishop, R. C. and T. A. Heberlein. 1979. "Measuring values of extra-market goods: Are indirect measures biased?" American Journal of Agricultural Economics, Vol. 6; 926-30.

Cooper, J. C. 1994. "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys." Land Economics, 70(1): 111-122.

Green, D., K. Jacowitz, D. Kahneman, and D. McFadden. Referendum contingent valuation, anchoring, and willingness to pay for public goods. Paper presented at the First Toulouse Conference on Environment and Resources Economics. Toulouse, France. March 30-31, 1995.

Greenley, D. A., R. G. Walsh, and R. A. Young. 1981. "Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality," Quarterly Journal of Economics, 96(Nov.): 657-672.

Hanemann, M. 1984. "Welfare evaluations in contingent valuation experiments with discrete responses," American Journal of Agricultural Economics, Vol. 66;335-79.

____________, 1989. "Welfare evaluations in contingent valuation experiments with discrete responses: Reply," American Journal of Agricultural Economics, Vol. 71:1056-1056.

Johansson, P., B. Kriström, and K. G. Mäler, 1989. "Welfare evaluations in contingent valuation with discrete responses: Comment," American Journal of Agricultural Economics, Vol. 71:1055-1056.

Keith, J.E., C. Fawson and V. Johnson. (Forthcoming). "Preservation or Use: A Contingent Valuation Study of Wilderness Designation in Utah." Ecological Economics.

Larsen, D.M. 1993. "On Measuring Existence Value," Land Economics, Vol 69: 377-388.

McConnell, K.E, 1983. "Existence and Bequest Values." In: Rowe, R. D. and L. G. Chestnut, Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas, Boulder, Colorado: Westview Press.

Pope, A. C. III and J. Jones. 1990. "Value of wilderness designation in Utah." Journal of Environmental Management, Vol. 30:157.

Smith, V. K. 1993. "Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal," Land Economics, vol. 69: 1-26.

United States Department of the Interior, Bureau of Land Management. Utah BLM Statewide Wilderness Final Environmental Impact Statement. Washington, D. C.: U.S. Government Printing Office. 1990.

Utah Wilderness Coalition. Wilderness at the Edge: A Citizen Proposal to Protect Utah's Canyons and Deserts. Salt Lake City, Utah: Utah Wilderness Coalition. 1990.

Walsh, R., J. Loomis, and R. Gillman, 1984. "Valuing Option, Existence, and Bequest Demands for Wilderness," Land Economics, Vol. 60: 14-29.

ENDNOTES

1

. There is at least one other significant alternative proposal, the Hansen-Orten option. However, this alternative had not been sufficiently defined at the time of the study to provide the study respondents with enough detailed information on which to compare the three proposals. In addition, there have been many less well specified proposals reported in the press and elsewhere.

2

.The map used was taken from Wilderness at the Edge.

3

.CV questions were also asked of opponents to wilderness in general, as well opponents of the BLM and UWC proposals. See Keith, Fawson and Johnson for a discussion of those responses.

4

. There has been considerable discussion of closed-ended, referendum CV questions in the literature. Some authors (Green, et al., 1995, for example) suggest that this approach causes overestimates due to anchoring effects compared to open ended questions. Others suggest that the close-ended approach is based on random utility functions, while the open ended approach has no such underlying consistency.

A CHARACTERISTICS APPROACH TO ESTIMATING THE VALUE OF ATLANTIC SALMON SPORTFISHING

by

Vishwanie Maharaj
American Sportfishing Association
1033 North Fairfax Street, Suite 200
Alexandria, Virginia 22314

James J. Opaluch
Environmental and Natural Resource Economics
University of Rhode Island
Kingston, Rhode Island 02881

June 1996

INTRODUCTION

This paper describes a study designed to assess the potential use of aquaculture for augmenting Atlantic salmon stocks for sport fishing purposes, and examines issues related to improving available methods of benefit transfer. The study uses a contingent behavior survey to estimate the value of Atlantic salmon sport fishing opportunities with varying attributes, including catch rate, fish size and wild versus stocked fish, among other attributes.

The value estimates are integrated with a model that determines the costs associated with raising and stocking fish of various ages to assess the economic feasibility of using pen-reared Atlantic salmon to augment natural populations. We also assess the use of different sets of variables of angler characteristics for transferring benefit measures to other sites. Specifically, we identify the relative explanatory powers of three sets of characteristics: demographic characteristics of anglers, measures of level of interest in fishing and motivations of anglers for fishing (e.g., catch-related motives versus desire to be natural areas).

The overall model is depicted in Figure 1. The ultimate goal of the study is to select stocking strategies that maximize net benefit, where various attributes of the fish are under the control of the decision maker. These attributes include the number of fish released, the life stage at which fish are released (juveniles versus adults) and the number of river-miles stocked. The left side of Figure 1 indicates the demand side of the model, where the number of anglers and the surplus per angler are determined. The right side of the diagram represents the supply side, where the number of fish and the unit cost of fish are determined. Along the bottom of the diagram in the of the model. This includes the characteristics of the site at which stocking occurs (e.g., developed areas near towns versus areas far from development), the allowable congestion on salmon rivers, the amount of fishable waters, target levels for catch rates, fish size, etc.

This paper discusses the demand side of the model only. First, we describe a conceptual model of recreational values and relate values to site characteristics and angler characteristics. We then describe

Figure 1

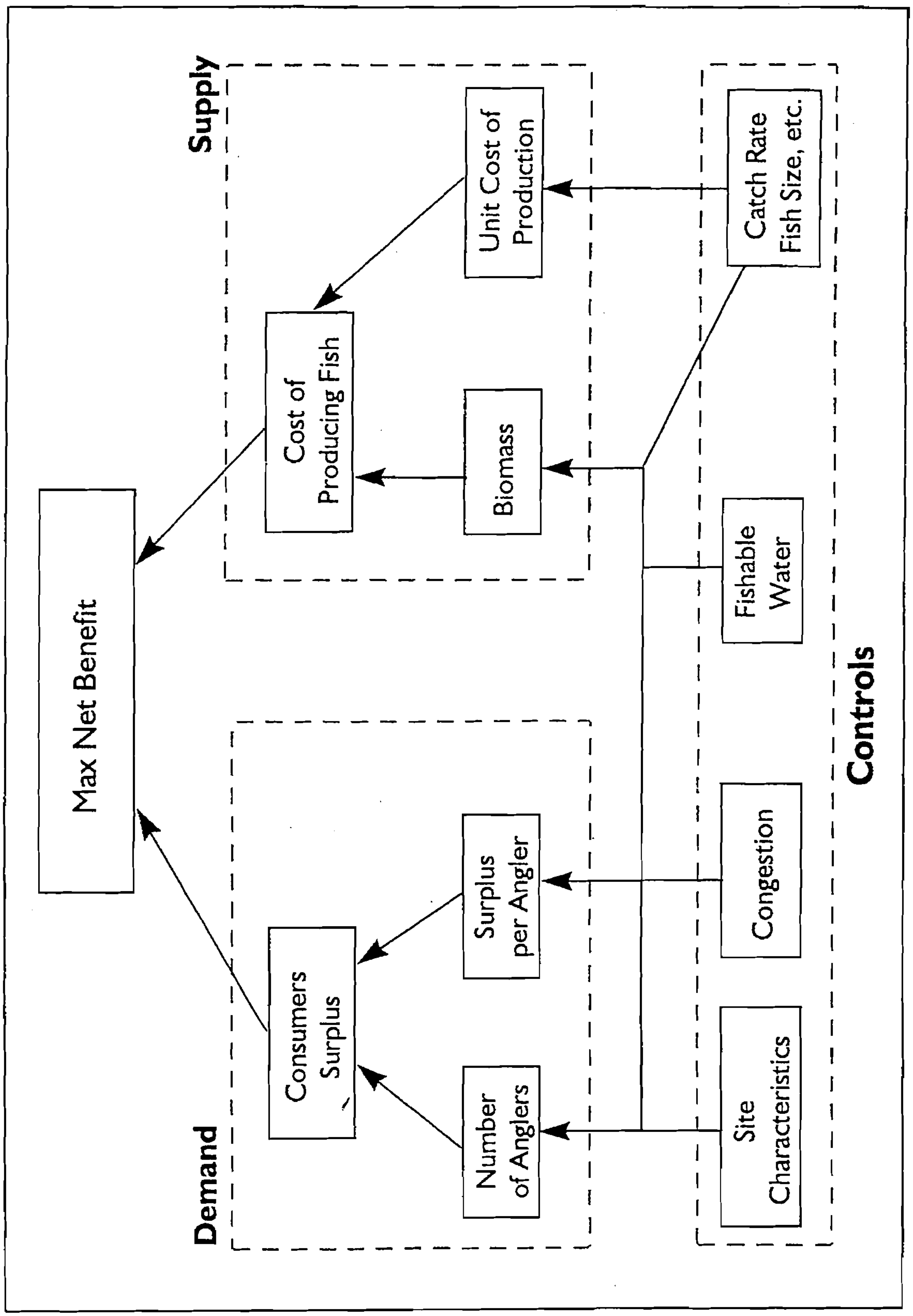


Figure 1 Model Depiction

a survey used to measure sportfishing values as a function of these characteristics and describe the empirical results of the survey. We also discuss issues regarding the explanatory power of different angler characteristics and the extent to which these sets of characteristics can potentially improve our ability to transfer benefit estimates between different contexts.

Preliminary results indicate a high value for Atlantic salmon sport fishing, as well as a considerable potential for aquaculture operations augmenting current fish populations in Maine. We found that demographic variables had very poor explanatory power in measuring differences in values across anglers, but that level of interest variables and angler motivations had high explanatory power. This indicates that we may be able to greatly improve our ability to transfer benefit estimates among different contexts if we could expand the information on participant characteristics that is collected in standard surveys, such as the NMFS fishing survey and the US Fish and Wildlife hunting and fishing survey.

THEORETICAL MODEL

The angler is assumed to face the following maximization problem:

$$\text{Max } U=[X, Q, T; C_X\, C_A)]$$

$$\text{subject to } I = P_X X + P_Q' Q + P_T' T$$

where X represents the number of days spent fishing at the hypothetical site, P_X is the cost of participation, Q is a vector of non-leisure goods, P_Q is a vector of prices for non-leisure goods, T is a vector of time spend in other recreation activities, P_T is a vector of prices of other recreation activities, I is income, C_X is a vector of attributes for the recreation activity and C_A is a set of angler characteristics.

Under suitable assumptions, the maximization problem can be solved for a system of Marshallian demand functions:

$$X = f(P_X, P_Q, P_T, I; C_X, C_A)$$

where demand at a fishing site depends on prices, site attributes (C_X), and angler attributes (C_A). If variable that explain consumer heterogeneity are not included in the estimated demand function, angler attributes become omitted variables that are part of the error term e_i:

$$X = f(P_X, P_Q, P_T, I; C_X, C_A) + e_i.$$

In contrast, if we can measure angler attributes that are important in determining differences in demand, and we can estimate the impact of these attributes on demand, then we can explain a larger portion of the variance. Thus, our ability to transfer benefit estimates might be greatly improved to the extent that we can identify important characteristics that determine demand at both the study site and the policy site.

In this study, differences in demand within angler populations (consumer heterogeneity) are modeled using three sets of variables. First, we use traditional demographic variables, like age, income and education. We also include a set of variables indicating the angler's level of interest in Atlantic salmon sport fishing and a set of variables indicating motives for fishing. We then test these three sets of variables for explanatory power, which would indicate the usefulness of each set for benefit transfer.

DESCRIPTION OF THE SURVEY

Survey development

The survey was developed through an extensive process which included ethnographic interviews (Spradley 1979), focus groups (Desvouges and Smith, 1988) and verbal protocol (Schkade and Payne, 1994) pretests of preliminary survey instruments. The survey development process was initiated in the Summer of 1992 and extended through spring of 1993.

The ethnographic interviews were carried out with 12 anglers from two sport fishing clubs in Maine during September of 1992. Ethnographic interviews are in depth one-on-one interviews, where the respondent is the "expert" who provides information. A key notion to the ethnographic method is that attributes used to distinguish different categories of a commodity are meaningful and important elements of the commodity to the respondent. Thus, the interviewer can identify aspects of the commodity that are important to respondents by getting them to describe different categories of the commodity.

For example, if one wanted to know how people think about automobiles and which attributes are most important, the interviewer would ask the respondent to describe different kinds of automobiles. One respondent might distinguish between luxury cars versus sports cars, while another might discuss expensive cars versus inexpensive cars. The ethnographic method would infer from this that "style" is an important element to the first respondent, while price is an important element to the latter. The ethnographic method would follow up on these comments by getting individuals to talk more about these elements, to identify why they are important, and to categorize automobiles in other ways. This allows the interviewer to understand which elements are most important, why they are important, how people think about automobiles, and what terms they use in thinking about automobiles, with minimal leading of respondents.

In applying the ethnographic method, we asked each angler to categorize and describe different kinds of fresh water sport fishing experiences. Respondents were asked to compare and contrast these different kinds of experiences, thereby revealing characteristics that determine preferences and influence choice behavior. Each angler was also asked to provide a range of values for these characteristics. This information was used to develop levels of characteristics for the hypothetical sport fishing scenarios in the survey. We also asked participants to describe different types of Atlantic salmon anglers in order to identify characteristics of anglers to be used in the analysis. These interviews were also used to identify terms used by anglers, which was important so that survey questions and information provided to survey respondents could be described in a meaningful way.

Following the initial one-on-one interviews, focus group meetings were held to discuss the information collected and to pretest survey questions. In initial survey questions, we used a small number of attributes in describing salmon fishing sites for fear that questions with many attributes could become too complex (Mazzotta and Opaluch, 1995). However, we found that respondents consistently asked for more information on hypothetical salmon sites. Salmon anglers in focus groups could handle quite complex survey questions, which included numerous attributes. Indeed, respondents were unwilling to make decisions when they did not have adequate information on attributes of fishing experiences. This appears to indicate that respondents with considerable decision making experience with respect to a commodity are able to formulate responses regarding quite complex commodities. The ultimate survey instrument included 10 site characteristics, as indicated in Table 1.

Survey Implementation

An in-person self administered survey of Atlantic salmon anglers was conducted during June and July, 1993. In Maine, rivers attract large numbers of salmon anglers. Only a certain number of anglers are allowed to fish at one time, and the remaining anglers wait on the side of the river for their turn to fish. The survey was carried on-site, with a self administered survey booklet given to anglers who were waiting to fish.

The survey instrument contains contingent behavior questions, where hypothetical Atlantic salmon sport fishing scenarios were used to elicit hypothetical demand (Maharaj, 1995). Respondents were given some preliminary information regarding the survey and were shown a poster size example question which was used to explain to respondents how to fill out the question. Respondents were presented with hypothetical salmon fishing experience, described in terms of the site characteristics. Respondents first asked to rate each attribute on a scale of 1 to 10. Respondents were then asked whether they would fish there if such a fishing site were available, and if so, how many days would they fish there. We asked respondents to rate attributes for two reasons. First, having respondents rate attributes would ensure that they paid close attention to each attribute. Secondly, attribute ratings will allow different methods for analyzing the data, based on ratings of individual attributes, as well as stated participation regarding hypothetical sites, which are bundles of attributes. To date, we have not utilized the attribute ratings.

Table 1 Site Characteristics in the Model

Site Attribute	Level
Site Type:	1. 20 Minutes from town, no visible development, good access roads 2. 20 Minutes from town, houses and a dam visible, good access roads 3. Fairly remote area, access roads not paved for 5 miles, camping at \$25/night 4. Fairly remote area, access roads not paved for 5 miles, lodging and meals at \$150/night
Pool Length:	Varied from 50 yards (3 "lies") to 300 yards (15 "lies")
Anglers in pool (Congestion)	1 to 9
Mode of fishing	1. Pools assessable from shore or by wading 2. Pools assessable by boat only 3. Boat rental available
Fish Type	1. Wild Fish 2. Fish stocked as fry or smolts 3. Fish stocked as adults 3 months prior to migration 4. Fish stocked as adults immediately prior to migration
Catch Rate:	Ranged from 3 fish per day to 1 fish in 4 days
Fish size:	Ranged from 7-9 lbs. to 13-17 lbs.
Season:	Spring (May, June, July) Fall (September and October)
Additional driving time to site:	Ranged from 1 to 6 Hours
Daily rod fee:	Ranged from \$50 to \$300

In addition to information on sites, we collected data on angler demographics, level of interest in Atlantic salmon sport fishing and motives for participation (Maharaj, 1995). The Angler characteristics upon which we collected data are presented in Table 2.

A Tobit model was used to estimate a hypothetical demand for salmon sport fishing. Catch rate, fish size, congestion, travel time, and daily rod fee are all highly significant and have expected signs. Results indicate that anglers are indifferent between catching fish stocked at the juvenile stages, and fish stocked as adults. However, these anglers expressed a higher level of satisfaction from catching wild fish as compared to stocked fish.

In the econometric models, a number of variables are used to represent the anglers level of interest in Atlantic salmon sport fishing as indicated in Table 2. The level of interest variables include the amount of money invested in Atlantic salmon fly fishing equipment (gear and boat) divided by income; the amount of money spent on Atlantic salmon fishing trips during the last year fished divided by the angler's total expenses on all vacation trips; and the number of days the angler went fishing for Atlantic salmon during the last year fished divided by total vacation time in that year. This resulting set of transformed variables are used as indicators of the level of interest of anglers who have different time and monetary budgets.

The other level of interest variables are dummy variables. Anglers' rating for this sport is collapsed into two categories: those who rated Atlantic salmon fishing as more enjoyable than other outdoor leisure activities and those who rated the sport as just as or less enjoyable than other outdoor leisure activities. Dummies were used to indicated anglers reported their skill levels in three categories: average, above average and novice.

Motives for Atlantic salmon fishing were accounted for using a Likert scale from 1 to 5, where anglers were asked to indicate the importance of various motives (See Table 2). For empirical analysis the sample is divided into two groups. One group is comprised of anglers who are motivated to fish for

Table 2. Angler Characteristics

Demographics		State of Residence Age Gender Education Head of household (Yes/No) Number of dependents Income Employment status
Motives (Likert Scale)	Non-Catch related	Mental relation Solitude See and learn about wildlife Be outdoors and explore natural surroundings Explore new fishing sites Fish close to home Be with Friends Family recreation
	Catch related	See others catch fish Catch fish which is challenge Catch fish that fights a lot Develop/share skills Catch large fish Catch wild fish Obtain fish for eating
Level of Interest		Amount spent on fishing Member of salmon club Number of articles Tie own flies Take time off from work Fish at least 4 times a week % of vacation time spent fishing Years spent fishing

Atlantic salmon because of catch-related reasons versus those whose non-catch related motives are just as or more important than their catch related motives.

Catch related motives are: to obtain fish for eating; to see others catch fish; to catch lots of fish; to catch wild fish; to develop and share skills; to catch large fish; to catch fish that are good fighters/challenging fish. Non-catch motives were: family recreation and other social reasons; being outdoors and exploring new sites; fishing close to home; to see and learn about wildlife; for solitude and mental relaxation. For each angler the average rating for all catch related motives (C) and the average rating for all non-catch related motives (NC) are calculated. A dummy variable is then created, which is set to 1 if C > NC and is zero otherwise. After this sample separation there are 188 anglers with catch related motives and 172 with non-catch related motives.

EMPIRICAL MODEL AND RESULTS

A large portion of respondents indicated responses to the number of days that they would fish at the hypothetical site, hence the Tobit model is used to analyze this censored data (Greene, 1990). Intercept surveys usually over sample frequent participants, as they are present at the interview site more often and have a higher likelihood of being selected (Nowell *et al,* 1988). This disproportionate representation can lead to serious estimation problems (Manski and Lerman, 1977). However, weighting these observations appropriately will correct for this sample selection bias, and yield consistent estimates (Nowell et al, 1988). Thus, weighted Tobit models are estimated using Limdep (Version 6). The weighted log likelihood function is:

$$\sum\nolimits_0 \ln(w_i(1-\Phi(\beta' Z_i/\sigma)) + \sum\nolimits_1 w_i \ln(2\pi\sigma^2)^{-1/2} - \sum\nolimits_1 w_i[(X_i - \beta' Z_i)^2/2\sigma^2)]$$

$$w_i = [N/\sum\nolimits_i W_i]^* W_i$$

where W is the weight specified. (Greene, 1991).

Several specifications of the Tobit demand model are estimated. In the initial model, it is postulated that the expected number of days an angler visits a site is a function of the attributes of the site only (Model 1). Tobit demand models with only site attributes as explanatory variables are modified to explore the influence of angler characteristics on sport fishing demand. Different models are specified to include demographics, variables reflecting the angler's level of interest in Atlantic salmon sport fishing, and motives for salmon fishing. Results from these models are then compared to determine the relative explanatory power of different sets of angler characteristics.

Estimation results are shown in Table 3. Model 1 includes only site characteristics. All coefficients are statistically significant and of the correct sign. Model 2 is constructed by adding demographic characteristics to model 1, resulting in an improvement in fit that is significant at the 95% level. However, only income is statistically significant, while age, education, and resident status are not statistically significant.

Table 3. Tests of Explanatory Power of Sets of Characteristics

	Likelihood Ratio Tests		Degrees of	Significance
Model	Models	Chi-Sq.	Freedom	Level
1. Site Char Only				
2. Demographics	1 vs. 2:	44.4	4	95%
3. Level of Interest	2 vs. 3:	200.6	3	99%
4. Motivation	3 vs. 4:	4.8	3	80%

Variables that reflect level of interest in this sport are included in model 3, together with site attributes and income. A likelihood ratio test overwhelmingly rejects the null hypothesis that the model with site attributes only gives as good a fit to the data as the model that includes level of interest variables. Coefficients on all level of interest variables have positive signs. Thus, demand will increase with stated level of interest. All level of interest variables are significant at least at the 95% level except for the ratio of fishing days to vacation days and the high rating for salmon fishing. These insignificant coefficients could have resulted from multi-collinearity as the condition number of the matrix of explanatory variables 78 (see, for example, Belsley Kuh and Welsch, 1980). Finally, model 4 includes all angler characteristics in model 3 plus angler motivation variables.

In the models estimated above it is assumed that the same value of the unknown parameters for site attributes are applicable to all members of a population. However, the motivation variable clearly show that anglers have different preferences for site attributes. Interactions between the motivation dummy variable (catch-related motivation versus non-catch) and the site attributes to capture these taste differences, which is equivalent to estimating separate coefficients on each site attribute for the two angler groups. Accounting for these taste differences improves the explanatory power of the level of interest model, as indicated in Table 3.

As expected, anglers whose primary motives are catch related have a higher value for catch related attributes of the sport fishing experience. Anglers whose non-catch related motives are relatively more important place less weight on catch related attributes. In Model 5 main effect coefficients represent the non-catch related group, while dummy interactive variables represent the difference between this group and the catch related angler group.

For anglers with non-catch motives, none of the coefficients on fish type are significantly different from the base case. Thus, for these anglers there is no discernible difference in preference for wild fish, fish stocked as smolts or fry, or fish stocked as adults. In contrast, fish type is very important to anglers with catch related motives as there is a preference ordering for fish type, where the coefficient on wild fish is higher than the coefficient on smolts, which in turn is higher than the coefficient on fish stocked as adults 3 months prior to migration.

The coefficient on fish size is insignificant, however the fish size dummy interacted with the dummy variable for catch-related motives is statistically significant. The catch rate variable interacted with the dummy variable indicating catch-related motives is also highly significant, indicating that the catch rate coefficient is significantly higher for anglers with catch motives, as expected.

IMPLICATIONS FOR BENEFITS TRANSFER

Cost and time savings can be significant if it is possible to transfer a model's parameters from one location to another, or from one time period to another. In situations where research cost exceeds the benefits of a primary study, it may be preferred to transfer available benefit estimates to that context or location. Even though benefits transfer is less costly than conducting original studies, it may produce less accurate results and should only be carried out if appropriate studies are available. Furthermore, most authors agree that demand or value functions as opposed to average unit values should be used for benefits transfer (Krupnick, 1993; Smith, 1993). Transfer functions could be developed from past studies using meta analysis and applied to other locations/contexts that are similar to those in the model (Smith and Kaoru, 1990). In addition, for meta analysis to be successful these studies must value a wide range of site and user characteristics (Walsh *et al*, 1989).

A major problem in benefits transfer could arise if populations at the study and policy sites do not share a common representative utility function. As a result, transfer of benefits from one group to another may result in large errors if the representative utility differs greatly across groups. Benefits transfer can be improved to the extent that these differences can be measured and explained, emphasizing the importance of including consumer heterogeneity in estimated models (Fletcher *et al*, 1992).

In order to illustrate the importance of correctly specifying benefits transfer functions, several demand transfers exercises are conducted, where Tobit demand models are estimated for the following subsets of this study's data:

1. All study data except for a subset of data from Eastern Maine.
2. All study data except a subset of data from Northern Maine.
3. All data from the Penobscot area.
4. All study data except for a group of anglers who have a high level of interest..
5. All study data except for a group of anglers who have catch related motives.

For each data set four specifications of the weighted Tobit demand model are estimated: a model with only site characteristics (Site model); a model with site attributes and demographics (Demographics model); a model with site attributes and variables reflecting level of interest in Atlantic salmon sport fishing (Level of interest model); and a model with site attributes, variables reflecting level of interest in Atlantic salmon sport fishing, and dummy interactive variables for anglers with catch related motives (Motives model). In the ensuing discussion, each of the five data sets used in model estimation are referred to as study samples, while those not used in model estimation are referred to as transfer samples.

If it is assumed that no information exists on any of the explanatory variables, including site characteristics, the only recourse is to transfer average values from the study site to the respective transfer site. In each of the following transfer exercises average demand for the study sample (naive estimate) is compared to the actual estimate for the transfer sample. This would indicate the order of magnitude of the error that could arise from transferring an average estimate to another location, without considering site conditions and user characteristics at that location.

Also, for each subset of the data, all four specifications of the Tobit model are applied to the respective transfer sample. Site models do not account for population characteristics but explain variations in demand resulting from differences in site attributes. Results from demographic models, level of interest models, and motives models account for user characteristics. Comparison of results from these latter three models indicate the explanatory power of each set of user characteristics.

Three measures are used to compare the performance or predictive ability of these models: average demand; correlation of actual and predicted demand; and the percentage of best predictions. Even though average demand estimates may be close to the true average demand, this should not be the only criterion used to measure a model's predictive ability. The mean of these estimates may mask wide deviations of individual predictions from the true value. A correlation coefficient is a good measure of the relationship or association between two sets of data (Judge et al. 1985). In addition, the percent predictions that are closest to the true value are compared for all four models. First, for each observation in the transfer sample, the absolute deviation of estimated demand from true demand is calculated for all models. Then predictions from all four models are ranked according to their deviation from the true value, where the estimate with the lowest deviation is ranked one and the estimate with the highest deviation is ranked four. The predictive ability of each model is compared using the percentage of estimates ranked 1 to 4.

In all cases, except for the transfer sample of anglers with catch motives, the naive estimate is the least accurate, as shown in Table 4. The widest difference among these models is observed for the transfer sample where anglers have a high level of interest. Compared to the true mean, average demand estimates from the demographics and site models have a much higher deviation than average estimates from the level of interest and motives models.

For all transfer samples, estimated demand from demographic models have a slightly higher correlation to true demand as compare to demand estimates from site models. However, level of interest models give much better results than demographic models. For all transfer samples except for the one with anglers who have catch motives, results from the motive models only produce a slight improvement in correlation over level of interest models. For this catch motives sample, the motives model did much better than the level of interest model.

Percentage of predictions ranked one and two show the same trend as the correlation results. Compared to demographics and site models, use of motives models and level of interest models produce more predictions ranked 1 or 2. These differences are widest for the transfer sample from Northern Maine, the catch motives transfer sample and the level of interest transfer sample.

After accounting for site attributes, it appears that demographics do not add much explanatory power to transferred estimates. Furthermore, any improvement can be fully attributed to the presence of income among these variables. Variables reflecting level of interest and motives for fishing are better at characterizing angler populations. Inclusion of these characteristics in a benefits transfer function is important, especially if the transfer population is very different from the "average" angler in the study

Table 4. Benefit Transfer Results

Study Data Set Models		Transfer Data Set Correlation Coefficients	Demand Estimates (Days per Angler)
Transfer From: Northern Maine & Penobscot		Transfer to: Eastern Maine	Actual Data (E. Maine) = 2.38 Naïve Transfer Estimate = 2.08
Models:	Site	.02	2.14
	Demographics	.04	2.17
	Level of Interest	.15	2.41
	Motives	.17	2.35
Transfer From: Eastern Maine & Penobscot		Transfer to: Northern Maine	Actual Data (N. Maine) = 2.12 Naïve Transfer Estimate = 1.78
Models:	Site	.18	1.89
	Demographics	.35	1.94
	Level of Interest	.46	2.13
	Motives	.49	2.21
Transfer From: Penobscot		Transfer to: Northern & Eastern Maine	Actual Data (N. & E. Maine) = 2.06 Naïve Transfer Estimate = 1.73
Models:	Site	.14	1.89
	Demographics	.14	1.94
	Level of Interest	.21	2.13
	Motives	.26	2.21
Transfer From: All Remaining Data		Transfer to: Random Selection of Data with High Level of Interest	Actual Data (High Level of Interest) = 3.70 Naïve Transfer Estimate = 1.89
Models:	Site	.37	1.94
	Demographics	.40	2.14
	Level of Interest	.52	3.17
	Motives	.57	3.00
Transfer From: All Remaining Data		Transfer to: Random Selection of Data with Catch-Related Motives	Actual Data (Catch-Related Motives) = 1.74 Naïve Transfer Estimate = 2.16
Models:	Site	.59	2.26
	Demographics	.67	2.35
	Level of Interest	.70	2.13
	Motives	.74	1.86

population. Thus, original valuation studies should measure not only site attributes but also characteristics that reflect level of interest and motives for participating in a sport. This will facilitate the transfer of benefits to another site/location or population of anglers.

SUMMARY AND CONCLUSIONS

This paper describes two components of a study which examined the potential economic feasibility of using aquacultured salmon to increase populations of Atlantic salmon for sportfishing. The paper describes a contingent behavior survey used to estimate consumer surplus values obtained from stocking fish and issues in transfer of these benefit estimates to other sites.

Tobit demand models are estimated, where expected demand is postulated to be a function of attributes of Atlantic salmon sport fishing, demographics, level of interest in this sport, and motives for going fishing. Income is the only statistically significant demographic variable. Education, age, and resident status are insignificant. In addition, the outcome of statistical tests indicate that demographic variables as a group do not add much explanatory power to the demand model. However, variables indicating the angler's level of interest in this sport are highly significant. Demand is higher for anglers who spend a large portion of their vacation budget on Atlantic salmon sport fishing, have a significant percentage of income invested in gear, are highly skilled, rate the sport as highly enjoyable, and spend a large amount of time salmon fishing given their availability of leisure time.

Motives for going salmon fishing are indicators of angler preferences for attributes of this sport. Catch rate, fish size, and fish type are important choice variables for anglers whose catch related motives are more important than their non-catch related motives. In contrast, these variables have less influence on the choice behavior of anglers whose non-catch related motives are relatively more important. Furthermore, anglers with catch motives clearly prefer wild fish over hatchery reared fish, even if the latter were stocked as juveniles (fry/parr/smolt). On the other hand, for anglers with non-catch motives there is no discernible difference between demand for sites with wild fish and sites with stocked fish. Tobit demand models are estimated for different sub sets of this data and used in demand transfer exercises. Models that include level of interest variables and motives for fishing give better out of sample predictions than models with only demographics and site attributes.

Given these results, it is recommended that original valuation studies should not only measure site attributes but also characteristics that reflect level of interest and motives for participating in a sport. These variables could be very useful for benefit transfer if similar characteristics were collected by various standard surveys, such as the NMFS recreational fishing survey and the U.S. Fish and Wildlife recreation survey. User characteristics are also be useful for government agencies for provide conditions that attract the type of participants agencies want to attract and they could be used to monitor shifts in demand over time as tastes and preferences. This would enable more accurate benefit transfer of studies in different time periods.

REFERENCES

Belsley, D.A., E. Kuh and R. E. Welsch, 1980. *Regression Diagnostics* New York: John Wiley and Sons.

Desvouges, William H. and V. Kerry Smith. Focus Groups and Risk Communication: The 'Science' of Listening to Data. *Risk Analysis*, 8, 479-484 (1988).

Federal Register 1994. NOAA proposed rules for damage assessment. Federal Register, 59(5): 1147-1150.

Fletcher, J. J., W. L. Adamowicz, and T. Graham-Tomasi. 1992 The travel cost model of recreation demand: Theoretical and empirical issues. *Leisure Sciences*, 12:119-147.

Greene, W. H. 1991. *Limdep Version 6. User's Manual and Reference Guide*. Econometrics Software Inc.

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

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl, and T. C. Lee. 1985. *Introduction to the Theory and Practice of Econometrics* (second edition). John Wiley and Sons, New York.

Kennedy, J. J. and G. Brown. 1976. Attitudes and behavior of fishermen in Utah's Uinta Primitive area. Fisheries, 1(6): 15-31.

Krupnick, A. J. 1993. Benefit transfers and valuation of environmental improvements. Resources for the Future, Winter No. 110: 1-7.

Kuentzel, W. F. and C. McDonald. 1992. Differential effects of past experience, commitment, and lifestyle dimensions on river use specialization. *Journal of Leisure Research*, 24(3): 269-287.

Maharaj, V. 1995. *Valuation of Atlantic Salmon Sportfishing in New England and an Economic Analysis of Farming Adult Atlantic Salmon for a Sport Fishery*. Ph.D. dissertation. University of Rhode Island, Kingston, RI 02881.

Manski, C. F. and S. R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples". *Econometrica*, 45 (8): 1977-1988.

Mazzotta, M. J. and J. J. Opaluch, 1995. "Decision Making When Choices Are Complex: A Test of Heiner=s Hypothesis" *Land Economics* Vol. 71 No. 4 (November).

Nowell, C., M. A. Evans, and L. McDonald. 1988. "Length-biased Sampling in Contingent Valuation Studies". *Land Economics*, 64(4):367-371.

Schkade and John Payne, 1994. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis" *Journal of Environment Economics and Management* Vol. 26 Pp 88-100.

Smith, V. K. 1993. "Nonmarket Valuation of Environmental Resources: An Interpretative Appraisal". *Land Economics*, 69(1): 1-26.

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

Walsh, R. G., D. M. Johnson, and J. R. McKean. 1989. "Issues in Nonmarket Valuation and Policy Application: A Retrospective Glance". *Western Journal of Agricultural Economics*, 14(1): 178-188.

Wessells, C. W., J. J. Opaluch, S. K. Swallow, and D. Wichelns. 1990. *A Landfill Site Evaluation Model that Addresses Public Preferences Regarding Impacts on Natural Resources and Local Communities*. A final report to the Rhode Island solid waste management corporation. Prepared by the Department of Resource Economics, University of Rhode Island.

REGIONAL HOUSEHOLD PREFERENCES FOR ECOSYSTEM RESTORATION AND SUSTAINED YIELD MANAGEMENT OF WILDERNESS AND OTHER NATURAL AREAS *

Richard G. Walsh
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, CO 80523

Ken C. Bonetti
Department of Economics
Colorado State University
Fort Collins, CO 80523

J. Michael Bowker
Southeastern Forest Experiment Station
Forest Service, U.S.D.A.
Athens, GA 30602

Kun H. John
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, CO 80523

Richard L. Johnson
National Biological Service
U.S. Department of Interior
4512 McMurray Avenue
Fort Collins, CO 80521

ABSTRACT

A national sample of 380 households was resurveyed in 1994 to test the possibility of a significant statistical relationship between willingness to pay and increments in the preservation of natural areas in five regions and the United States. The hypothesis of a significant relationship, consistent with the theory of diminishing marginal benefit, is supported in this case. The study is an experiment in cost-effective consumer research to demonstrate how future inquiry into the subject of household preferences and use could contribute to measuring the geographic distribution of benefits from ecosystem restoration and sustained wildlife habitat management. Also, this study contributes to the process of developing an understanding of other significant variables that may predict changes in regional benefit estimates, including income, distance, simple reminders, quality of ecosystem services, direct use for outdoor recreation trips, and indirect use for indoor recreation activities such as reading, watching programs, and hearing about the subject. The research objective is benefit transfer, that is, comparing past and current studies to estimate benefits for long-run policy analysis.

* This study was funded, in part, by cooperative research agreement No. 29-978 between the USDA-Forest Service, Southeastern Forest Experiment Station, Outdoor Recreation and Wilderness Assessment Research Unit, Athens, Georgia, 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 several individuals especially Julie Schaefers, Wayde Morse, John McKean, and John Loomis.

INTRODUCTION

Publicly owned natural areas that remain roadless and undeveloped are estimated to contain about 360 million acres or 15 percent of total land usage in the United States. Yet only about a quarter of this land has been subject to restoration or protected under sustained management in designated wilderness, parks, recreation areas, wildlife refuges, nature preserves, and the like. The map and chart in Figure 1 illustrate how much natural area remains and the amount that is protected in five regions. Current natural areas are approximately 5 percent of total regional land use in the North, compared to the South, 4 percent; Rocky Mountains, 15 percent; Pacific Coast, 25 percent; and Alaska, 50 percent. Protected natural areas are estimated as 1 percent of total regional land use in the North and South; compared to the Rocky Mountains, 3 percent; Pacific Coast, 5 percent; and Alaska, 15 percent.

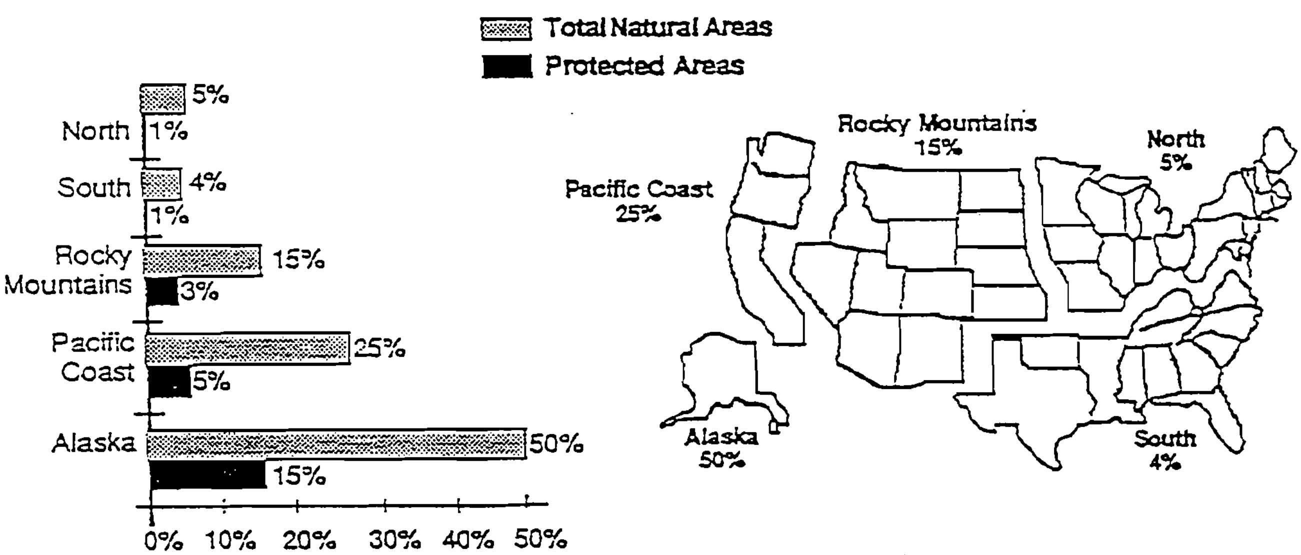


Figure 1. Amount of Regional Natural Area Protected as Wilderness, Parks, Wildlife Refuges, etc., United States, 1994

The importance of the economic value of natural areas is particularly evident where they have alternative uses for agriculture, timber harvest, mining, energy, and water development. According to a recent study, at least 1.5 million acres of natural areas change to other uses each year (English et al. 1993). If present trends continue, nearly one-third of remaining unprotected natural areas could change to other uses by the year 2040. The increasing scarcity of natural ecosystems makes it critical, in a balanced approach, to assess how their preservation also would contribute to national economic development.[1] Nations around the world face similar problems of estimating how much they can afford to pay for the protection of parks, wildlife refuges and other natural areas. The possibility of expanded development for other uses and the accompanying probability of damage to quality of ecosystems provides a realistic setting for investigating the significance of ecosystem restoration and sustained ecosystem management to the taxpaying public.

Natural areas are defined as undeveloped public land and water that provide fish and wildlife habitat, biological and genetic diversity, recharge of clean water and air, and other important environmental services. Natural areas usually do not have roads, buildings, mines, clear cuts, or related development. Natural areas conserve the ecosystem of soil, water, grassland, mountains, trees, other plants, fish and wildlife species, etc. People enjoy seeing a variety of wildlife, hiking, picnicking, camping, riding horseback, fishing, hunting, cross-country skiing, taking pictures, and other nonmotorized on-site activities. People also enjoy reading, watching programs, and hearing about natural areas.

The dual objective of ecosystem management, to preserve unique natural areas and make them available for the enjoyment of people, requires information on the benefits of both objectives (Roggenbuck and Walsh, 1993). For example, according to the U.S. Congress (PL 88-577, 1964), the objectives of wilderness designation is to protect the existence of natural ecosystems of plants, wildlife, and land forms

---

[1] Meadows suggests: "The necessary debate here is about how much nature to leave alone. Ten percent? (In many of our ecosystems, from tall-grass prairie to old-growth forest, it's too late for that.) Five percent? Two percent? We will have to stop eating into nature when we come to zero; there are...reasons to stop long before that."

for their own sake, while guaranteeing right of access to people under condition of sustainability; that is, the natural ecosystems remain unimpaired for future generations. Most state and federal wildlife and park agencies have similar objectives. The current draft 50-year management plan of the Forest Service emphasizes this dual objective in ecosystem restoration and sustained yield management of the National Forests (Thomas, 1995). This is a controversial subject about which very little is known according to a recent review by Resources for the Future (Sedjo, 1995).

The purpose of this study is to contribute to the best practicable application of economics to the valuation of ecosystem management objectives. It attempts to develop and apply cost-effective research procedures from successful consumer market survey methods to estimate household benefit functions for preservation of natural areas wherever they remain throughout the country. The primary objective is to test the possibility of a significant relationship between willingness to pay and the amount of natural area protected as wilderness, parks, etc. in five regions and the U.S. It is hypothesized that household demand for natural area preservation is consistent with the theory of diminishing marginal returns. Previous regional studies in two Rocky Mountain states and a Canadian Province suggest that willingness to pay may be a function of increments in amount of wilderness protected. The null hypothesis of no significant relationship between willingness to pay and the scale of natural area preservation, also has received support owing to an imbedding effect for three wilderness sites located in the Rocky Mountains.

The secondary objectives of this study is to contribute to the process of developing an understanding of other significant variables that may predict changes in benefit estimates, including income, distance, simple reminders, quality of ecosystem services, direct use for outdoor recreation trips, and indirect use for indoor recreation activities such as reading, watching programs, and hearing about the subject. The overall research objective is benefit transfer, that is, comparing past and current studies to estimate benefits for long-run policy analysis.

REVIEW OF LITERATURE

The economic valuation of wilderness and other natural areas has traditionally focused on the demand for onsite recreation use.[2] Fewer studies have included economic valuation of the preservation objective of management based on contingent valuation of willingness to pay.[3] While the present study includes the demand for onsite recreation use and preservation values, it differs from earlier work by introducing household economic valuation of increments in the protection of natural areas in the five regions of the United States. This would complement previous research on trends in recreation use. The future demand for protection and recreation use of natural areas is forecast to far outpace supply in all regions, particularly the Rocky Mountains and Pacific Coast, with a shortage of 50-60 percent projected by the year 2040 (English, et al. 1993). Direct use is related primarily to ease of access or travel distance, while indirect use is expected to be more related to quality of attributes and services of natural areas, less constrained by distance from place of residence.

---

[2] These studies include the recreation use value of the Pemegewasset Wilderness in New Hampshire (Halstead, et al. 1991); Ramseys Draft Wilderness in Virginia (Prince, 1988); Boundary Waters Wilderness in Minnesota (Walsh, et al. 1989); Linville Gorge Primitive Area in North Carolina (Leuschner, et al. 1987); Ventana Wilderness in California (Smith and Kopp, 1980); Glacier Park, Goat Rocks, Diamond Peak, and Eagle Cap Wilderness in Oregon and Washington (Brown and Plummer, 1979); two primitive areas in Utah (Loomis, 1979, 1980); Washakie Wilderness in Wyoming (Barrick and Beazley, 1990); the Indian Peaks, Comanche Peak and Rawah Wilderness in Colorado (Rosenthal and Walsh, 1986; Walsh and Gilliam, 1982); 10 million acres of wilderness and other natural areas in Colorado (Walsh, et al. 1984); 11 potential wild and scenic rivers in Colorado (Sanders, et al. 1991); 15 potential wild and scenic rivers in Alabama (Clonts and Malone, 1990); and the Colorado River in the Grand Canyon (Bishop, et al. 1988). In addition, a review of 120 studies completed over a 20-year period (Walsh, et al. 1990) reveals that there are many more studies of activities such as deer and elk hunting, trout fishing, and whitewater boating whose quality is enhanced by the protection of natural areas as wilderness, wildlife refuges, wild and scenic rivers, etc. (Loomis, 1992).

[3] These studies include the coastal wilderness and other natural areas on Prince William Sound, Alaska (Carson, et al. 1992); the Lye Brook, Big Branch, Breadloaf, George D. Aiken, and Bristol Cliffs Wilderness in Vermont (Gilbert, et al. 1992); the Washakie Wilderness in Wyoming (Barrick and Beazley, 1990; Diamond, et al. 1993); the Selway Bitterroot Wilderness in Idaho and the Bob Marshall Wilderness in Montana (Diamond, et al. 1993); 16.2 million acres of wilderness and other natural areas in Utah (Pope and Jones, 1988); 10.0 million acres of wilderness and other natural areas in Colorado (Walsh, et al. 1984; Aiken, 1985); 13.0 million acres of wilderness in the northern Rocky Mountains of Colorado, Idaho, Montana, and Wyoming (Diamond, et al. 1993; McFadden, 1994); to increase designated wilderness from 5% to 10% and 15% of the land base in British Columbia, Canada (Reid, et al. 1995); 11 potential wild and scenic rivers in Colorado (Aiken, 1985; Sanders, et al. 1990); 15 potential wild and scenic rivers in Alabama (Clonts and Malone, 1990); and the Colorado River in the Grand Canyon (Welsh, et al. 1995).

Table 1 summarizes the existing research on regional household willingness to pay for preservation of natural areas as wilderness in North America. The five studies report local household willingness to pay for protection of the resource in the region where they live. The studies of one to four states represent the Rocky Mountain region of the U.S., plus a province in western Canada. Three of the case studies suggest that local willingness to pay may be a function of increments in amount of the resource protected in each region, consistent with the economic theory of diminishing marginal utility. However, one study also asked separate samples of regional residents to value protection of one to three specific wilderness sites (Diamond, et al. 1993). The site specific approach appears limited by the problem of imbedding, where respondents tend to value a single site as not significantly different from the value of several sites in the region. At the 1994 meeting of W-133, Richard Carson argued that the three sites seem clustered too near the vertical axis to enable the researchers to reject the possibility of estimating a statistical demand curve for wilderness from contingent valuation research.

The contingent valuation results of the five case studies suggest that most local residents favor preservation of natural areas in the vicinity of where they live and would be willing to pay for it. The estimated values are likely to be conservative because the relevant population is limited to regional residents. Not included are possible benefits to tourists and the general public who do not visit the region, which results in the "aggregation" problem discussed in guidelines to nonmarket valuation research (Mitchell and Carson, 1989). This paper summarizes an attempt to begin evaluating other regions and to estimate national benefits. While people throughout the nation may have less interest in the subject, participate less in nature-related recreation activities, and express lower willingness to pay per household, when spread over the larger population, total public benefits of natural areas in each of the regions may be much greater than for local residents.

Table 1. Regional Studies of Household Willingness to Pay for Protection of Natural Areas as Wilderness in North America, 1994 Dollars.

Study Area, Resource, and Source	Total Quantity, Acres and Percent of Total Land Usage	Average Annual Willingness to Pay per Household	Investment Value per Acre at 6% Interest[a]	Year, Sample, and Population	Type of Survey and Valuation Format
COLORADO	Million Acres				
Roadless underdeveloped natural area protection as wilderness. Walsh, Loomis, and Gilman (1981, 1984).	1.2 million (2%) 2.6 million (4%) 5.0 million (8%) 10.0 million (15%)	\$21 \$30 \$43 \$63	\$395 \$260 \$194 \$142	1980, 195 households in state with 1,354,000 households (1980)	Mail survey, CVM open-ended questions, WTP into public special fund, TCM use value.
COLORADO Roadless underdeveloped natural area protection as wilderness. Aiken (1985); Walsh (1986).	10 million acres (15%)	\$78	\$176	1983, 198 households in Fort Collins vicinity to represent state with 1,354,000 households	Personal interviews, CVM open-ended and interative questions. WTP added taxes and higher prices.
UTAH					
Roadless underdeveloped natural area protection as wilderness. Pope and Jones (1990).	2.7 million (5%) 5.4 million (10%) 8.1 million (15%) 16.2 million (30%)	\$68 \$82 \$96 \$118	\$267 \$161 \$126 \$77	1986, 291 households in state with 636,000 households	Phone survey, CVM open-ended questions, WTP into public special fund.
BRITISH COLUMBIA	Current				
Roadless underdeveloped natural area protection as wilderness in British Columbia province in western Canada. Reid, Stone and Whitley (1995)	11.8 million (5%) Double by 11.8 million (10%) Triple by 23.6 million (15%)	 \$95 \$122	 \$172 \$110	1993, 1,561 households in province with 1,278,000 households	Mail survey, CVM open-ended question. WTP added taxes and fees in compensation.
	Direct use value	\$150	\$91		
NORTHERN ROCKY MOUNTAINS	Current				
Fifty-seven currently designated wilderness areas in states of CO, ID, MT, and WY. Diamond, Hausman, Leonard, and Denning (1993); McFadden (1994)	13.8 million (5%) Selway Billerroot (ID) Bob Marshall (MT) Washakie (WY) Three Areas	\$81 \$51 \$38 \$31 \$46	\$235	1991, 1,229 households in region with 2,267,000 households	Phone interviews, CVM open-ended and dichotomous choice questions. WTP added federal income taxes.

[a] 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.

RESEARCH PROCEDURE

Data for the study are from a resurvey of 380 households by a national market research firm in 1994. The useable response rate was 74 percent of the sample of 512 households who replied to a previous survey. The sample frame was a consumer panel stratified to represent U.S. Census household characteristics within each geographic region according to city size, annual household income, and size of household, as illustrated in Table 2. 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 national market research firm participates in surveys for the National Travel Data Center, National Conference Board, and major corporations. The basic panel sample stratification by region may be a more cost-effective way (at $12 per case) to represent the U.S. regional population than the alternative of conducting personal interviews with random cluster samples from a few representative communities, as in the Alaska state oil spill study (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 applied social science research.

Possible bias introduced by over or under sampling can be reduced by substituting the correct sample proportion in a statistical regression, unless the households sampled have more interest in the subject than the population (Mitchell and Carson, 1989). The potential problem is that those who respond to a survey may be more interested in and have higher values even if they do not differ demographically from those that do not respond. Then, simply reweighting the sample observations to give greater importance to under-represented sample groups would not fully correct for the problem of nonresponse. The study includes 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).

Table 2. Comparison of the Characteristics of Sample Households to Total Census Households, United States, 1994

Characteristics of Households[a]	Total U.S. Household, percent	Total Sample, Percent		Characteristics of Households	Total U.S. Households, percent	Total Sample, Percent	
		Selected Sample, (1,000)	Surveys Returned, (512)			Selected Sample, (1,000)	Surveys Returned, (512)
Geographic Region				***Household Size***			
New England	5.3	5.1	6.1	1 member	25.9	25.7	27.5
Middle Atlantic	14.9	15.0	15.4	2 members	32.0	32.1	36.3
East North Central	17.2	17.1	18.9	3 members	16.7	16.8	16.4
West North Central	7.4	7.6	8.0	4 members	15.3	15.3	13.5
South Atlantic	18.1	18.1	19.5	5 or more members	10.1	10.1	6.3
East South Central	6.2	6.3	5.3	Average number of members	2.6	2.5	2.4
West South Central	10.5	10.5	9.0				
Rocky Mountains	5.5	5.3	5.1	***Household Income***			
Pacific Coast	14.9	15.0	12.7	Under $12,500	18.3	17.9	17.6
Size of City				$12,500 to $24,999	20.9	20.8	20.7
Under 100,000	22.0	22.2	22.7	$25,000 to $39,999	21.8	22.2	20.9
100,000 - 499,999	16.4	16.4	18.8	$40,000 to $59,000	19.3	19.3	21.1
500,000 - 1,999,999	20.6	20.9	19.1	$60,000 and over	19.7	19.8	19.7
2,000,000 and over	41.0	40.5	39.5	Average Inc. ($1,000s)	$40.2[b]	$39.2	$40.1

a From the Current Population Survey, March 1992 (machine-readable data file)/conducted by the Bureau of the Census, Bureau of Labor Statistics — Washington: Bureau of the Census, 1992. U.S. Bureau of the Census, Current Population Reports; Series P-20, No. 467, "Household and Family Characteristics, March 1992."

b Assumes household money income reported by Census as $37,922 for 1991 increased by 6 percent to 1994.

Ten percent of the nonrespondents were interviewed by phone and a sample of 137 U.S. households were interviewed by random digit dialing. Chi-square comparisons of the three samples in Table 3 show that there is a significant difference in the proportion who say they have seen a natural area and their expressed interest in the subject; however, contrary to expectations, sample mean responses to the phone surveys are significantly higher than the mail survey. This suggests the mail sample may not over-state public interest in the subject.

Questions were designed for clarity and ease of answering (Dillman, 1978). Alternative questions were pretested on three samples of 25-100 persons. The questionnaire was printed on good quality paper, photo-reproduced, visually uncluttered, and bound in booklet form (see the Appendix). A map and artistic

Table 3. Comparison of Mail and Phone Surveys of Household Interest in Direct and Indirect Use of Wilderness and Other Natural Areas, United States, 1994

Direct and Indirect Use	Respondents to Mail Survey N = 512	Phone Survey, 10% of Nonrespondents N = 50	Random Digit Dialing, Phone Survey N = 137	Chi-Square Test of Significant Difference[a]
Have ever seen on trips				
YES	65.2%	78.0%	88.3%	25.39
NO	31.3%	22.0%	10.9%	0.01
No answer	3.5%	0.0%	0.7%	
Total	100.0%	100.0%	100.0%	
Interest in seeing on trips				
YES, Interested in seeing, total	78.3%	92.0%	92.7%	178.13
Very interested	45.9%	58.0%	67.9%	0.01
Somewhat interested	32.4%	34.0%	24.8%	
NO, Not interested in seeing	14.3%	8.0%	5.8%	
No answer	7.4%	0.0%	1.5%	
Total	100.0%	100.0%	100.0%	
Ever seen in any media				
YES	64.6%	86.0%	89.1%	
NO	34.8%	14.0%	10.2%	
No answer	0.6%	0.0%	0.7%	
Total	100.0%	100.0%	100.0%	
Newspapers				
YES	33.6%	50.0%	52.6%	29.60
NO	65.8%	50.0%	36.5%	0.01
No answer	0.6%	0.0%	10.9%	
Total	100.0%	100.0%	100.0%	
Magazines				
YES	49.1%	66.0%	67.9%	31.56
NO	50.3%	34.0%	21.2%	0.01
No answer	0.6%	0.0%	10.9%	
Total	100.0%	100.0%	100.0%	
Books				
YES	30.7%	50.0%	46.0%	25.94
NO	68.8%	50.0%	43.1%	0.01
No answer	0.6%	0.0%	10.9%	
Total	100.0%	100.0%	100.0%	
Television/Videos				
YES	52.4%	78.0%	85.4%	87.26
NO	47.0%	22.0%	3.6%	0.01
No answer	0.6%	0.0%	10.9%	
Total	100.0%	100.0%	100.0%	
Movies				
YES	23.8%	48.0%	46.0%	43.29
NO	75.6%	52.0%	43.1%	0.01
No answer	0.6%	0.0%	10.9%	
Total	100.0%	100.0%	100.0%	

[a] The first value is the chi-square test statistic and the second value is the probability that a phone survey is different (visibly higher) than the mail survey. Cases with no answer were omitted from the comparison. Other tests (likelihood ratio and Mantel-Haenszel) had similar results.

reproductions of the resource were shown on the questionnaire and on the cover letter. The term, natural area, was chosen over the more precise, ecological system, for clarity of understanding by the general reader. The payment vehicle was reduction of household income. The modified payment card format was based on successful commercial market survey methods, where respondents are asked to check a preprinted amount or write in any other amount. The letter was addressed to each individual by name, signed by the firm's familiar project leader, and was designed to motivate respondents by explaining the usefulness of the research to recreation resource planning and the importance of participating in the study. The sponsoring agencies were not identified to avoid possibly influencing respondents.

HOW MUCH TO PROTECT

Respondents were asked how much natural area they believe should be preserved in the region where they live and in other regions. They were offered a range of amounts including zero, 25%, 50%, 75%, 100% of the existing amount, and asked to check the amount they believe should be preserved or to write in any amount, including an added amount by restoration. This approach was designed to provide respondents with sufficient information to understand the concept of supplying incremental quantities and substitution, thus reducing possible imbedding effects of the single-site approach. The difference is analogous to the single-site vs. multi-site approach recommended in applications of the travel cost method.

Table 4 shows that if residents have an opportunity to choose size of the preservation program, the average amount preferred is about 94 percent of the existing amount.[4] The 95-percent confidence interval is 89-98 percent. An 86.3 percent majority of households say they want to preserve a positive amount, 2.1 percent do not want any, and 11.6 percent are undecided or did not answer the question. The regional frequency distributions show a majority prefer preservation of 100 percent of the existing amount. The

[4] With the information on alternative supply available, fewer respondents were undecided or did not answer and the 95-percent confidence interval around the mean value was narrower than for the willingness to pay question.

Table 4. **Regional Household Expressed Preferences and Willingness to Pay for Preservation of Natural Areas in Regions of the United States, 1994**

Expressed Preference and Willingness to Pay for Regional Resources[a]	North Residents N=149		South Residents N=125		West Residents N=91		United States N=380	
	Average	Standard Error	Average	Standard Error	Average	Standard Error	Average	Standard Error
Amount of Preservation								
North	101.4%	5.5	102.9%	7.9	91.5%	7.5	99.1%	3.8
South	106.5%	7.3	109.5%	8.4	99.9%	8.7	105.4%	4.6
Rocky Mountains	92.3%	2.8	92.1%	3.2	89.7%	4.4	91.5%	1.9
Pacific Coast	87.5%	2.7	88.7%	3.2	85.0%	4.2	87.4%	1.8
Alaska	90.4%	2.2	89.9%	3.1	83.1%	3.7	88.5%	1.7
United States	94.5%	3.4	96.7%	4.4	88.1%	5.2	93.5%	2.4
Willingness to Pay								
North	$54.06	6.1	$63.22	13.0	$40.70	10.4	$53.05	5.6
South	$38.08	5.3	$87.62	13.1	$30.46	6.9	$52.59	5.4
Rocky Mountains	$42.95	6.1	$55.91	8.1	$54.74	10.9	$49.10	4.4
Pacific Coast	$36.92	5.4	$49.15	8.2	$65.48	14.3	$46.81	4.7
Alaska	$57.46	11.2	$63.91	9.6	$73.50	16.6	$61.74	6.6
United States, Total	$229.47	28.7	$319.82	37.4	$264.88	46.2	$263.28	20.1

[a] For cases reporting zero or positive willingness to pay. About 10 percent of the values are constrained to a maximum of $1,000 per household for the U.S.

Questions: **How much** natural areas do you believe should be preserved in the region where you live and in other regions of the U.S.? **(Check ONE Box For EACH Region); How Much in EACH Region?** None, 25% of Existing Amount, 50% of Existing Amount, 75% of Existing Amount, 100% of Existing Amount, More Than 100% Through Restoration (Write in), and Not Sure.

This question is hypothetical and intended to provide an economic measure of how much these sites (reported in Question above) 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 them from changing to other uses. Consider your household income and other things you could purchase with the money **(Check One Amount For EACH Region)**; $0, $1, $5, $10, $20, $30, $40, $50, $60, $70, $80, $90, $100, $200, $300, $400, $500, $750, $1,000, Other (Specify), and Not Sure.

frequency distributions average: 25 percent (7.1%), 50 percent (7.9%), 75 percent (7.1%), 100 percent (54.7%), and more than 100 by restoration (9.7%).

Preference for preservation of regional resources ranges from 88 percent of existing natural area in Pacific Coast states and Alaska to all of the existing amount in Northern states and more than the existing amount in Southern states. Residents of states in the North and South tend to prefer restoration of more natural areas in the Eastern regions where they live than do residents of states in other parts of the U.S. Residents of the South favor protection of the existing amount and restoration of 10 percent more natural area in the region, including the deteriorated Everglades in Florida. Residents of the North favor restoration of 6 percent more in the South.

The proportion of households reporting a preference for more than 100 percent of the existing amount of wilderness and other natural areas through restoration averages: U.S., 0.10 percent; North, 0.15 percent; South, 0.17 percent; Rocky Mountains, 0.10 percent; Pacific Coast, 0.07 percent; and Alaska, 0.04 percent. For these cases, the amount preferred averages: U.S., 168 percent; North, 193 percent; South, 202 percent; Rocky Mountains, 147 percent; Pacific Coast, 152 percent; and Alaska, 145 percent.

Apparently, natural areas are viewed as national assets wherever they remain throughout the U.S. For example, residents of the West say they prefer preservation of a larger proportion of existing natural area in the South (99.9%) and North (91.5%) than in the West (Rocky Mountains, 89.7%; Pacific Coast, 85.0%; and Alaska, 83.1%). Also, residents of the North and South favor preservation of somewhat more natural areas in the West (88.7-92.3%) than do residents of the West (83.1-89.7%).

WILLINGNESS TO PAY

Respondents were asked how much the preferred amount of natural area in each region is worth to them. They were reminded that this is the maximum amount of money they would pay per year to protect it, and asked to assume that payment of a part of their income is the only way to prevent natural areas changing to other uses.[5] Respondents were offered a range of values and asked to check the highest amount they would be willing to pay rather than forego preservation of the natural areas they desire. The range includes zero and 18 increasing values from $1 to $1,000, a place to write in another value, and an undecided category.

Table 4 shows that households are willing to pay an average of $263 per year for preservation of 94 percent of existing natural areas in the five regions. The 95-percent confidence interval is $243 to $303. A 56.7 percent majority of the households say they are willing to pay a positive amount, 8.3 percent are not willing to pay,[6] 28.7 percent are undecided, and 6.3 percent did not answer the hypothetical question intended to provide an economic measure of how much natural areas are worth to them. The regional frequency distributions show a solid core of values with a peak of about $50, and then tail off to an upper value of $1,000. The frequency distributions average: $1 (2.7%), $5-10 (16.3%), $20-50 (19.6%), $60-100 (9.2%), $200-750 (8.5%), and $1,000 (0.2%).

The preservation value for natural areas in the total U.S. is significantly higher for residents of the South ($320) than residents of other parts of the nation ($229-$265), even though household income in the South is substantially less, averaging $35,800 compared to $44,350 in the North and $41,580 in the West. Over all households sampled, the most valuable natural areas are located in Alaska ($62). Residents of the North value natural areas in their own regions ($54) second only to Alaska resources ($57). The same is

[5] In addition, one-half of the households were reminded to consider their household income and other things they could purchase with the money. The other one-half of the sample were asked the identical willingness to pay question without the reminder. The test of the NOAA proposal that willingness to pay estimates without an income and substitute reminder be decreased by 50 percent is not sustained in this case. The direction of the effect lends support to the principle of adjustment, although for the small sample, the means with and without the added information are not significantly different similar to Loomis, et al. (1994). While the income and substitution reminder, willingness to pay averages $247 (SE=27) compared to $280 (SE=30) without the reminder, a decrease of $33 or 12 percent. It is noteworthy that with the reminder, fewer households refuse to answer (3.8% vs. 8.4%) and more report zero value (10.7% vs. 6.4%).

[6] Included are approximately 2.7 percent whose response to a reasons question indicate rejection of the payment vehicle of a portion of their income.

true for residents of the West who value natural areas in their own region ($55-$65) second only to Alaska resources ($74).

Residents of the South value natural areas in their own region ($88) substantially more than in Alaska ($64). Residents of other regions value natural areas in the South less than residents of the South, with the difference significant at the 5 percent level for residents of the North ($38) and at the 10 percent level for residents of the West ($30), based on Steel and Torrie (1980, 173). Also, residents of the West value natural areas in the North ($41) less than in the Rocky Mountains ($55) and Pacific Coast ($65), although the difference is not statistically significant for the small sample (91 cases).

REGIONAL HOUSEHOLD BENEFIT FUNCTIONS

Table 5 describes six equations estimating the statistical relationship between willingness to pay and the amount of natural area preservation. In the five regional functions, willingness to pay in a region is the dependent variable and the amount of natural area preservation in the region is the independent variable. In the sixth function, the dependent variable is the sum of willingness to pay in all five regions and the independent variable is the sum of natural area preservation. These equations assume that household preference for amount of natural area preservation is an ex post or exogenous past decision that conditions ex ante or endogenous future willingness to pay similar to other endowments, i.e., education, income, leisure time, etc. which are fixed.[7] The adjusted R^2 ranges from 22-32 percent, and the F-statistic from 21.3-51.5 which indicates that the equations are highly significant. The standard errors shown in parentheses indicate the linear, quadratic and cubic variables are significant at the 0.01 level.

[7] Similar results may be obtained by assuming that both decisions are endogenous, and applying a two-stage multiple regression model. To assume exogenous quantity facilitated use of a simple regression model, which was effective for a small sample with wide variation in household willingness to pay. The regional benefit functions presented here should be considered tentative approximations for illustrative purposes, subject to revision with further study.

Table 5. Household Willingness to Pay for Increments in the Preservation of Wilderness and Other Natural Areas in Regions of the United States, 1994[a]

Region	Average Willingness to Pay	Average Quantity Preference	Linear Quantity	Quantity Squared	Quantity Cubed	Adjusted R^2	F-statistic
North N =227	$39.80 (5.35)	102.1 (8.2)	0.920886* (0.16)	-0.005751* (0.002)	8.01902E-06* (2.6313E-06)	0.22	21.8
South N = 220	$40.90 (5.47)	110.6 (9.4)	0.742957* (0.12)	-0.003633* (0.000)	4.21129E-06* (1.4615E-06)	0.22	21.3
Rocky Mountains N = 221	$42.03 (4.85)	93.2 (6.8)	0.945157* (0.13)	-0.005014* (0.001)	-	0.32	51.5
Pacific Coast N = 208	$39.55 (4.83)	89.0 (6.5)	0.944825* (0.15)	-0.005228* (0.001)	-	0.31	47.7
Alaska N = 210	$57.45 (7.21)	89.9 (6.5)	1.345134* (0.26)	-0.007203* (-0.002)	-	0.31	48.2
United States N = 242	$222.81 (27.21)	94.1 (36.4)	4.898959* (0.82)	-0.026829* (0.006)	12.2303E-06* (4.0928E-06)	0.28	37.9

* Significant at 0.01; standard errors in parentheses. For cases reporting willingness to pay a positive amount.

In Table 6, the statistical estimates of the relationships are used to calculate household willingness to pay for incremental levels of natural area preservation. The first line for each region and the U.S. shows the proportion preferred. The household values in the statistical functions are adjusted for the proportion of sample households reporting preference greater than or equal to increments in the preservation of natural areas. The second line shows the total value reported by households. The values plot out Bradford-type (1970) public benefit functions where willingness to pay is a function of increments in natural area protection as wilderness in each region. The functions increase at a decreasing rate with the protection of additional natural areas. The third line shows the marginal value or change in total benefits resulting from one unit changes in the amount of natural area protected. The first derivatives or slope of the total benefit functions represent the demand curves for preservation of natural areas.

Table 6. Household Preference and Willingness to Pay for Preservation of Increments in Wilderness and Other Natural Areas in Regions of the United States, 1994

Household Values for Regions[a]	Number of Cases	None, Zero Intercept	Proportion of Existing Amount			
			25 Percent	50 Percent	75 Percent	100 Percent
North						
Proportion Preferred	318	0.01	0.99	0.86	0.78	0.73
Total Value	261	0	$17.01	$28.42	$34.89	$37.06
Marginal Value, cents		80.1¢	56.4¢	35.3¢	16.9¢	01.0¢
South						
Proportion Preferred	308	0.01	0.99	0.90	0.80	0.75
Total Value	251	0	$14.90	$26.02	$33.72	$39.38
Marginal Value, cents		67.6¢	52.8¢	37.4¢	24.5¢	13.0¢
Rocky Mountains						
Proportion Preferred	309	0.01	0.99	0.91	0.82	0.74
Total Value	240	0	$18.04	$30.56	$37.58	$39.05
Marginal Value, cents		83.2¢	61.1¢	39.1¢	17.0¢	-5.0¢
Pacific Coast						
Proportion Preferred	307	0.03	0.97	0.90	0.80	0.70
Total Value	243	0	$17.50	$29.39	$35.64	$36.29
Marginal Value, cents		81.3¢	58.8¢	45.3¢	13.8¢	-8.7¢
Alaska						
Proportion Preferred	300	0.03	0.97	0.93	0.83	0.72
Total Value	242	0	$25.34	$42.85	$52.53	$54.39
Marginal Value, cents		117.0¢	85.7¢	54.4¢	23.1¢	-8.3¢
United States						
Proportion Preferred	336	0.02	0.98	0.90	0.81	0.73
Total Value	275	0	$92.79	$157.24	$194.36	$206.17
Marginal Value, cents		429.2¢	313.8¢	211.5¢	95.3¢	-8.0¢

[a] Household values from the quadratic and cubic functions (Table 5) adjusted for proportion of households reporting preference greater than or equal to increments in preservation and proportion willing to pay: North, 0.87; South, 0.91; Rocky Mountains, 0.87; Pacific Coast, 0.86; and Alaska, 0.87.

Figure 2 illustrates the household willingness to pay functions for alternative levels of natural area preservation in the five regions. Willingness to pay increases at a decreasing rate with the protection of additional natural area. For the existing natural areas protected as wilderness, wildlife refuges, and parks in each region, households are willing to pay a great deal because of its scarcity value. However, as more natural areas are designated for protection, the willingness to pay for each additional area becomes smaller, consistent with the theory of diminishing marginal benefits. As household demand for preservation of natural areas in each of the regions becomes fully satisfied, at about 90 to 110 percent of the existing amount, willingness to pay reaches a maximum. Beyond the optimal level, total value would diminish with further expansion of ecosystem restoration and sustained management.

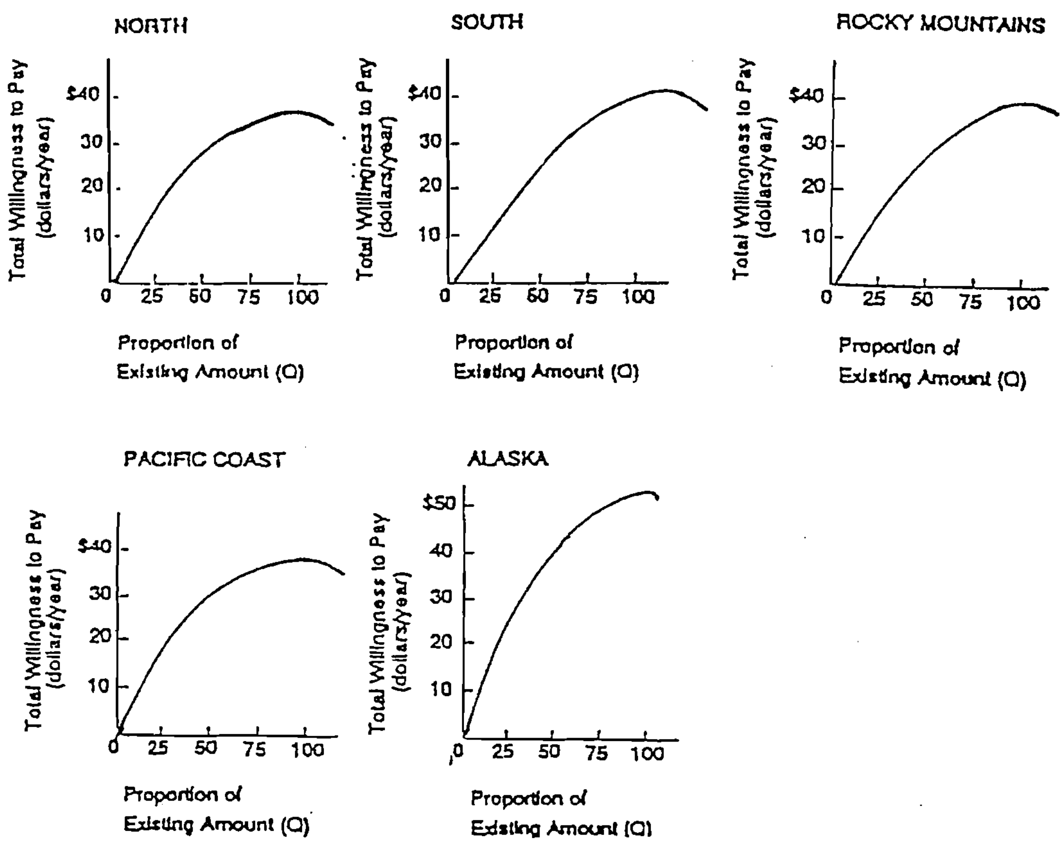


Figure 2. Annual Household Benefit Functions for Increments in Preservation of Natural Areas in Regions of the United States, 1994

Table 6 also shows the sum of household willingness to pay for the five regions. Total household willingness to pay increases at a decreasing rate from the origin at zero to $93 for 25 percent, $157 for 50 percent, $194 for 75 percent, to $206 for 100 percent of existing natural areas in the U.S. The contingent valuation estimate of household willingness to pay for natural area preservation ranges from the statistically adjusted \$206 to a reported average of \$263 per year for about 94 percent of the existing amount. An approximation of the present value of natural areas is the sum of household benefit each year aggregated over 96 million households and discounted at 6.0 percent. Present value of annual benefit would represent the investment value of the resource to households in the U.S. If natural areas are found on approximately 360 million acres representing 15 percent of land usage, the investment value to U.S. households would range from \$966 to \$1,241 per acre in 1994 dollars.

EFFECT OF SOCIOECONOMIC VARIABLES

Table 7 describes five regional equations which estimate the statistical relationship between willingness to pay and quality, distance, use, income and other socioeconomic variables. Adjusted R^2 ranges from .26 to .36, and the F-statistic from 9.3 to 10.4, indicating that the semilog equations are significant. The t-ratios shown in parentheses indicate variables are significant at 0.05 level or better, with two exceptions at 0.10. The results are preliminary based on a trimmed data set, with the number of cases reduced by nonreporting of some variables. This results in somewhat higher means for willingness to pay, income, participation, etc. than for the total sample.

For the most part, the contingent valuation results suggest that the determinants of nonmarket demand for public ecological services are similar to market demand for private goods and services in the household production of recreation activities. The dependent variable is the natural log of reported willingness to pay for preservation of the preferred amount of natural area in each region. Explanatory variables that are significant and positive include: perceived quality of the resource; participation in direct use on trips; indirect use reading, watching programs and hearing about the subject; reported benefits of time in these nature-related activities; and household income per member. The proxy for distance, resident of the East, is positive in regressions for the North and South, indicating willingness to pay for natural areas in these regions by residents of the West is negatively related to distance. The variable also is positive in the regression for Alaska, indicating that willingness to pay for natural areas in the region is

Table 7. **Preliminary Estimates of the Effect of Socioeconomic Variables on Willingness to Pay for Protection of Wilderness and Other Natural Areas, in Regions of the United States, 1994.**

	North		South		Rocky Mountain		Pacific Coast		Alaska	
Variable	Mean	Coefficient	Mean	Coefficient	Mean	Coefficient	Mean	Coefficient	Mean	Coefficient
Household Income per Member, Thousand Dollars	22.07	0.0406 (2.52)[a]	21.22	0.0469 (2.66)	22.34	0.0460 (2.51)	22.68	0.0451 (2.41)	23.02	0.0383 (2.23)
Quality of Natural Areas in Region, 1-5 Scale	3.77	1.1908 (3.74)	3.85	0.9168 (2.65)	3.96	1.9240 (4.36)	3.81	1.6168 (3.98)	4.12	1.3265 (3.22)
Have Seen Natural Areas, Binary, 0-1	0.94	2.6448 (1.75)	0.94	5.9670 (4.03)	0.94	4.5407 (2.51)	0.94	4.6971 (2.51)	0.94	3.7605 (2.20)
Indirect Use of Natural Areas, Binary, 0-1	0.94	6.2062 (4.28)	0.95	4.4425 (2.69)	0.96	6.7784 (3.01)	0.96	6.8841 (2.96)	0.96	7.3099 (3.51)
Reported Benefits of Natural Area Related Activities, Dollars per Hour	20.16	0.0278 (2.45)	21.04	0.0375 (3.30)	22.25	0.0423 (3.38)	21.98	0.0227 (1.76)	21.15	0.0361 (2.98)
Resident of East Region, Binary, 0-1	0.55[b]	1.9268 (2.97)	0.82	2.0102 (2.24)	--	--	--	--	0.79	1.8239 (2.08)
Age of Respondent, Years	47.72	-0.0759 (-3.43)	47.18	-0.0674 (-2.83)	46.90	-0.0741 (-2.63)	--	--	47.18	-0.0662 (-2.51)
Highest Education of Household Head, Years	--	--	15.03	-0.4010 (-2.40)	--	--	--	--	--	--
Unemployed Respondent, Binary, 0-1	0.06	-4.0532 (-2.86)	--	--	--	--	--	--	0.05	-5.1836 (-2.90)
Constant		-9.2357 (-4.03)		-5.3067 (-1.48)		-15.0040 (-4.97)		-16.9020 (-6.20)		-13.5420 (-4.76)
Number of Cases		162		147		139		137		134
Adjusted R^2		0.30		0.31		0.30		0.26		0.36
F-Statistic		9.43		9.26		11.02		10.40		10.27
Mean of Dependent Variable[c]		84.07		86.67		82.30		80.68		114.54

[a]T-ratios in parentheses. [b]North region. [c]Dependent variable is the natural log of reported willingness to pay for protection of preferred natural areas in the region. Number of cases reduced because of nonreporting some variables. For expansion to the total sample, multiply the means of the dependent variables as follows: North, 0.62; South, 0.61; Rocky Mountains, 0.60; Pacific Coast, 0.58; and Alaska, 0.54.

positively related to distance. The effect of distance may be inter-related with quality. Variables that are significant and negative include: unemployment status, age, and education of respondent in the South.

The largest effect of quality on willingness to pay is in the Rocky Mountains followed by the Pacific Coast, Alaska, North, and South. For the most part, this is related to perception of relative quality in these regions. The effect of direct use is greater in the South than other regions, consistent with the fact that outdoor recreation is year around. The South is followed by the Pacific Coast, Rocky Mountains, North, and Alaska. The largest effect of indirect use is for the most distant region Alaska, followed by the Pacific Coast, Rocky Mountains, North, and South. The reported benefits of time in these nature-related activities is greatest for the Rocky Mountains, followed by the South, Alaska, North, and Pacific Coast. There is little or no apparent difference in the effect of income on willingness to pay for regional natural area preservation. Other significant socioeconomic variables, such as age of respondent, show similar effects across regions. When quantity of natural area protection is included as an explanatory variable, the income variable is unchanged, while the other socioeconomic variables become insignificant and R^2 decreases.

Several additional tables available from the authors provide regional values for ecological services, direct recreation use on trips, and indirect recreation use. Some important implications of the data are discussed in the following three sections.

QUALITY OF REGIONAL ECOLOGICAL SERVICES

Respondents were asked to give their opinions about the quality of natural areas in the region where they live and other regions. They rated the relative quality of 13 attributes and services on a 5-point scale, with (1) very low quality, (2) low, (3) medium, (4) high, and (5) very high quality. The average scores were estimated along with standard errors.

People believe Alaska has the highest quality natural areas by far. For 12 of the 13 attributes, households rate the quality of Alaska resources significantly higher than other regions. The single exception is with respect to convenient location and accessibility (2.98), for which all other regions rate higher. Alaska quality is highest in: providing scenic beauty of a natural landscape unaltered by man (4.31); protecting rare and endangered species (4.28); knowing that future generations will have natural areas (4.21); protecting air and water quality (4.17); knowing natural areas exist for their own sake (4.12); knowing that in the future they have the option to go there if they choose (4.02); conserving natural areas for education and scientific study (4.00); preserving unique plant and animal ecosystems and genetic diversity (3.96); providing uncrowded hiking, camping, fishing, hunting, wildlife viewing, etc. (3.88); providing jobs and income from the tourist industry (3.69); and providing spiritual inspiration (3.42).

The next highest quality natural areas are located in the Rocky Mountains. Quality of attributes and services in the region are a distant second compared to Alaska, except with respect to knowing that in the future they have the option to go there if they choose. Quality of attributes and services in the Pacific Coastal states is a close third to the Rocky Mountains, except with respect to air quality and providing uncrowded hiking, camping, fishing, hunting, wildlife viewing, etc. for which the Rocky Mountains is noticeably superior. The two regions are considered equal with respect to convenient location and accessibility.

People believe the North and South have the lowest quality natural areas in the U.S. The South has slightly higher quality than the North with respect to all attributes except protecting water quality and conserving natural areas for educational and scientific study. Both areas tend to be ranked lower than Alaska, Rocky Mountains, and the Pacific Coast regions except with respect to convenient location and accessibility, for which the South is highest rated of all regions and the North is second, reflecting the geographic distribution of the population. Both regions are rated slightly higher quality than the Pacific Coast with respect to providing uncrowded hiking, camping, fishing, hunting, wildlife viewing, etc. The South is rated higher quality than the Pacific Coast with respect to knowing that in the future they have the option to go there if they choose, and knowing that future generations will have natural areas.

There is little or no evidence that people rate the quality of natural areas in the region where they live higher than in other regions. Residents of the North and South tend to rate the quality of natural areas in all five regions somewhat higher in quality than residents of the Rocky Mountains and Pacific Coast states. However, the central tendency is for the relative quality ratings of the five regions to be approximately the same for households throughout the U.S.

DIRECT RECREATION USE ON TRIPS

Respondents were asked to indicate if they have ever seen natural areas in the region where they live and in other regions; how interested they would be in doing so in the future; and to write in the number of miles they would be willing to travel.

Approximately two-thirds (67%) of the households report they have taken trips to see natural areas in regions of the U.S. Nearly two-thirds of the households report they have seen natural areas in the North (67%) and South (65%). About one-half have seen natural areas in the Rocky Mountains (53%) and Pacific Coast (46%). Very few households report they have seen natural areas in Alaska (10%). This is identical to the findings of the Alaska state oil spill study (Carson, et al. 1992), where 10 percent of a national sample of 1,043 households interviewed in their homes said they had seen natural areas in Alaska.

As expected, residents of each region are more likely to have seen natural areas in the regions where they live than in other regions. Most residents of states in the North have seen natural areas located in the North (86%). Residents of states in the South have seen natural areas in the South (79%). Residents of states in the West have seen natural areas in the West (80%). Also a surprising number of resident households in each region report they have seen natural areas in other regions of the continental United States. For example, residents of the North report they have seen natural areas in the South (62%), Rocky Mountains (49%), and Pacific Coast (33%). Most westerners have seen natural areas in the North (64%) and South (51%).

It is noteworthy that between eight and nine households in ten express an interest in future direct use on trips to see natural areas in each region. Apparently, people are attracted by the unique characteristics of regional natural resources. Interest in future trips may reflect the basic human interest in variety of experience expected in visits to various types of natural environment throughout the nation. Interest in future recreation trips to see natural areas in regions is consistently greater than indicated by actual trips taken in the past. Expressed interest in seeing natural areas in the future located in Alaska (82%) and the North (80%), exceeds past trips to Alaska (10%) and to the North (67%).

Households express more interest in seeing natural areas in the regions where they live than in other regions of the country, however, the difference is not statistically significant. Nearly nine out of ten of households living in the East express an interest in seeing natural areas in the East (North, 88%; South, 90%) and an even higher proportion of households living in the West are interested in seeing natural areas there (Rocky Mountain, 94%; Pacific Coast, 93%), although the East and West are not significantly different in this respect. Residents of the East (North and South) are nearly as interested in seeing natural areas in the West (Rocky Mountains, 84-89%; Pacific Coast, 81-83%; and Alaska, 82-83%) as natural areas in the East (North, 88-90% and South, 81-91%). Residents of the West are nearly as interested in seeing natural areas in the East (North, 89% and South, 84%) as natural areas in the West (Rocky Mountains, 93%; Pacific Coast, 93%; and Alaska, 83%).

INDIRECT RECREATION USE

Respondents were asked if they have ever read, watched programs, or heard about natural areas in the region where they live and in other regions; how interested they would be in doing so in the future; and to write in the number of hours per year they typically do so.

Approximately eight out of ten (79-85%) households say they have participated in these indirect recreation uses of natural areas in the five regions. As expected, residents of states in each region are more likely to report indirect use of natural areas in the regions where they live than other regions. But overall, noticeably more people have read, watched programs or heard about natural areas in Alaska and the Rocky Mountains, consistent with the unique quality in these regions.

Interest is even higher in future indirect recreation use of natural areas. Nine out of ten (91%) households say they are interested in participating in these recreation activities in the future. It is noteworthy that expressed interest is virtually identical for natural areas located in all parts of the country. Slightly more residents of southern and western states are interested in future opportunities for indirect use of natural areas in the regions where they live than in other regions. However, residents of northern states are just as interested in natural areas in other regions as in the region where they live.

On average, households report they devote about 113.7 hours per year (2 hours per week) to indirect recreation use of natural areas in the five regions. (This compares to a reported 145.7 hours purchasing clothing.) Residents of southern states devote more time (130.9 hours) to nature-related indoor recreation activities than people who live in northern or western states (100.4-103.5 hours). It is noteworthy that the households throughout the country devote a considerable amount of time to indirect use of natural areas located in the North (25.1 hours) and South (24.0 hours), but less than to those located in the Rocky Mountains (21.8 hours), Pacific Coast (21.8 hours), and Alaska (21.3 hours).

People living in the North and South tend to devote less time to indirect use of natural areas in the regions where they live than to those located in other regions. For example, residents of northern states report an average of 27.8 hours of indirect use of natural areas located in the region where they live; and they report an average of 20.6 hours, or 74.1 percent as much time is devoted to indirect use of natural areas in Alaska, plus 19.4 hours for the Rocky Mountains, 18.3 hours for the Pacific Coast, and 17.4 hours for the South. However, people living in the West tend to devote more time to indirect use of natural areas in the Western Region than the North and South. Residents of western states report an average of 21.1 hours of indirect use of natural areas located in the Rocky Mountain region plus virtually an identical amount of time devoted to indirect use of natural areas located in Alaska (21.2 hours), and even more time devoted to indirect use of natural areas located in Pacific Coast states (23.9 hours). Westerners report 18.7 hours for the North and 15.4 hours for the South.

SUMMARY AND CONCLUSION

The primary contribution of this paper was to illustrate possible household willingness to pay functions for alternative levels of natural area preservation in five regions and the U.S. The regional analysis supports the hypothesis of a significant functional relationship between willingness to pay and the amount of natural area protected, consistent with the theory of diminishing marginal utility. The null hypothesis of no significant relationship was rejected for all regions, notably the Rocky Mountains where past research on 1-3 sites had lent some support to the null hypothesis. For the few existing natural areas preserved as wilderness, wildlife refuges, parks, etc., in each region, U.S. households are willing to pay a great deal, \$1,400 - \$1,900 per acre, because of uniqueness and scarcity value. However, as more natural areas are designated for preservation, the willingness to pay for each additional area becomes smaller, indicating diminishing marginal benefits. This is consistent with the findings of past studies of household values in states and a Canadian Province. As household demand for restoration and preservation of natural areas in each of the regions becomes fully satisfied, tentatively estimated at about 90 to 110 percent of the existing amount, willingness to pay reaches a maximum of \$970 - \$1,240 per acre. Beyond the optimal level, total value would diminish with further expansion of ecosystem restoration and sustained yield management.

The regional benefit functions presented here should be considered tentative and subject to revision with further study. The national data suggests that limiting the sample to residents of a single state or region results in a significant understatement of the contribution of natural area preservation to the welfare of households throughout the country. The national data on household preference suggest that a majority would support preservation of about four times more of the existing natural area as wilderness, wildlife refuges, parks, etc. than current programs provide in all regions, including the smaller less spectacular

Southern and Eastern wilderness areas. Also, the national sample willingness to pay for preservation is about four times more than local residents, \$966-\$1,241 vs. \$110-\$235 per acre, for around 10 million acres in Colorado, Utah, British Columbia, and the Northern Rocky Mountains (Colorado, Idaho, Montana, and Wyoming). The economic value of the resource depends on household willingness to pay, amount protected, direct and indirect use, attractiveness or quality, available substitutes, interest rate, and most important, number of households in the study area. The willingness to pay functions presented here illustrate the significance of several possible variables causing the range in values among regions.

A potentially useful approach to the benefit transfer problem in the future would be to collect regional data from a much larger sample and apply multiple regression analysis, similar to regional travel cost demand models (Loomis, et al. 1995). If the basic model specification is reasonably complete, that is, if it includes the relevant explanatory variables in the correct functional form, then it could explain the variability in benefits embodied in differences among the explanatory variables. The net benefit estimate for a site lacking data would then be predicted by inserting appropriate values of explanatory variables into the model fitted to regional data. This approach could add important new information on preservation values to complement regional variation in estimates of resource benefits from previous meta-analysis of 120 site-specific studies of recreation use (Walsh, et al. 1990).[8]

[8] The review of outdoor recreation literature during the previous 20 years for the 1990 RPA reported four of the nine Forest Service regional variables had negative coefficients, indicating that consumer surplus was \$9-13 per day lower than the national average in the South, Pacific Coast, and Intermountain regions. This means that the Northeast, Northcentral, and Rocky Mountain regions had significantly higher consumer surplus per day of outdoor recreation.

REFERENCES

Aiken, Richard. 1985. "Public Benefits of Environmental Protection in Colorado." Master's Thesis, Colorado State University, Fort Collins.

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.

Barrick, Kenneth A., and Ronald L. Beazley. 1990. "Magnitude and Distribution of Option Value for the Washakie Wilderness, Northwest Wyoming, USA." Environmental Management 14:367-380.

Bishop, Richard C., Kevin J. Boyle, Michael P. Welsh, Robert M. Baumgartner, and Pamela Rathburn. 1986. Glen Canyon Dam Releases and Downstream Recreation: An Analysis of User Preferences and Economic Values. Final Report to the Bureau of Reclamation. HBRS, Madison, WI.

Bockstael, N. E., and McConnell, K. E. 1981. "Theory and Estimation of the Household Production Function for Wildlife Recreation." Journal of Environmental Economics and Management 8:199-214.

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.

Bowker, J. Michael, and John R. Stoll. 1988. "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." American Journal of Agricultural Economics 70:372-381.

Bradford, David F. 1970. "Benefit-Cost Analysis and Demand Curves for Public Goods," Kyklos 23(4):775-791.

Brown, Gardner, and Mark Plummer. 1979. Recreation Valuation. An Economic Analysis of Nontimber Uses of Forestland in the Pacific Northwest, Appendix A.5. Forest Policy Project, Washington State University, Pullman.

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.

Clonts, Howard A., and Joy W. Malone. 1990. "Preservation Attitudes and Consumer Surplus in Free-Flowing Rivers," in Social Science and Natural Resource Recreation Management, ed. Joanne Vining. Boulder, CO: Westview Press.

Cordell, H. K., and Bergstrom, J. C. 1991. "A Methodology for Assessing National Outdoor Recreation Demand and Supply Trends." Leisure Sciences 13:1-20.

Diamond, Peter A., Jerry A. Hausman, Gregory K. Leonard, and Mike A. Dening. 1993. "Does Contingent Valuation Measure Preferences?: Experimental Evidence." Contingent Valuation: A Critical Assessment. Amsterdam: Elsevier. pp. 41-85.

Dillman, Don A. 1978. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley & Sons.

Duffield, John W., Thomas C. Brown, and Stewart D. Allen. 1994. Economic Value of Instream Flow in Montana's Big Hole and Bitterroot Rivers. Research Paper RM-317, Rocky Mountain Forest and Range Experiment Station, U.S. Forest Service, Fort Collins, CO.

English, Donald B. K., Carter J. Betz, J. Mark Young, John C. Bergstrom, and H. Ken Cordell. 1993. Regional Demand and Supply Projections for Outdoor Recreation. General Technical Report RM-230, Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture, Fort Collins, CO.

Gilbert, Alphonse H., R. Glass, and T. More. 1992. in "Valuation of Eastern Wilderness: Extramarket Measures of Public Support." The Economic Value of Wilderness Proceedings of the Conference. Claire Payne, J. Michael Bowker, and Patrick Reed (eds.). Gen. Tech. Rpt. SE-78, Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. p. 57-70.

Halstead, John M., Brice E. Lindsay, and Cindy M. Brown. 1991. "Use of the Tobit Model in Contingent Valuation: Experimental Evidence from the Pemegewasset Wilderness Area." Journal of Environmental Management 33:79-89.

Hoehn, John P. 1991. "Valuing the Multidimensional Impacts of Environmental Policy: Theory and Methods." American Journal of Agricultural Economics 73:289-299.

Larson, Douglas M. 1993. "On Measuring Existence Value." Land Economics 69(4):377-388.

Leuschner, William A., Phillip S. Cook, Joseph W. Roggenbuck, and Richard G. Oderwald. 1987. "A Comparative Analysis for Wilderness User Fee Policy." Journal of Leisure Research 19(2):101-114.

Loomis, John B. 1992. "Importance of Joint Benefits of Wilderness in Calculating Wilderness Recreation Benefits," in The Economic Value of Wilderness. Claire Payne, J. Michael Bowker, and Patrick Reed (eds.). Gen. Tech. Rpt. SE-78. Asheville, NC: U.S. Dept. Of Agriculture, Forest Service, Southeastern Forest Experiment Station. p. 17-26.

Loomis, John B. 1980. Monetizing Equity-efficiency Trade-offs in Rationing Wilderness Use. Western Energy and Land Use Team, U.S. Dept. of the Interior, Fish and Wildlife Service, Fort Collins, CO.

Loomis, John B. 1979. Estimation of Recreational Benefits from Grand Gulch Primitive Area. Moab District Office, U.S. Dept. of Interior, Bureau of Land Management, Moab, Utah.

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.

Loomis, John B., Brian Roach, Frank Ward, and Richard Ready. 1995. "Testing Transferability of Recreation Demand Models Across Regions: A Study of Corps of Engineer Reservoirs," Water Resources Research 31(3):721-730.

Loomis, John B., and Richard G. Walsh. 1992. "Future Economic Value of Wilderness." in The Economic Value of Wilderness, Claire Payne, J. Michael Bowker, and Patrick Reed (eds.). Gen. Tech. Rpt. SE-78, Asheville, NC: U.S. Dept. of Agriculture, Forest Service, Southeastern Forest Experiment Station. p. 81-90.

McFadden, Daniel. 1994. "Contingent Valuation and Social Choice." American Journal of Agricultural Economics 76:689-708.

Meadows, Donella. 1996. "What Would Happen If We Went Beyond Partisan Politics." The Ecological Economics Bulletin 1(1):35-36.

Ministry of Forests. 1995. Wilderness Issues in British Columbia: Results of a 1993 Province-wide Survey of Households. Working Paper WP-6-010, BC Ministry of Forests and BC Ministry of Environment, Lands & Parks, Victoria, BC, Canada.

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

Pope, C. Arden, and Jeffrey W. Jones. 1990. "Value of Wilderness Designation in Utah." American Journal of Environmental Management 30:157-174.

Prince, Raymond, and Ehsan Ahmed. 1988. "Estimating Individual Recreation Benefits Under Congestion and Uncertainty." Journal of Leisure Research 20(4):61-76.

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.

Roggenbuck, Joseph W., and Richard G. Walsh. 1993. "Wilderness Visitor Management, Benefits, and Non-Use Values." Managing the National Wilderness Preservation System: Research Needs. National Wilderness Research Committee, Wilderness Management Working Group. Society of American Foresters, University of Idaho, Moscow, ID.

Rosenthal, Donald H., and Richard G. Walsh. 1986. "Hiking Values and the Recreation Opportunity Spectrum." Forest Science 32(2):405-415.

Sanders, Larry D., Richard G. Walsh, and John B. Loomis. 1990. "Toward Empirical Estimation of the Total Value of Protecting Rivers." Water Resources Research 26:1345-1357.

Sanders, Larry D., Richard G. Walsh, and John R. McKean. 1991. "Comparable Estimates of the Recreational Value of Rivers." Water Resources Research 27(7):1387-1394.

Sedjo, Roger A. 1995. "Ecosystem Management: An Uncharted Path for Public Forests." Resources, Fall(121):10,18-20.

Smith, V. Kerry, and Raymond J. Kopp. 1980. "The Spatial Limits of the Travel Cost Recreational Demand Models." Land Economics 56:64-71.

Steel, Robert G. D., and James H. Torrie. 1980. Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill, New York.

Sutherland, Ronald J., and Richard G. Walsh. 1985. "Effect of Distance on the Preservation Value of Water Quality." Land Economics 61(3):281-291.

Thomas, Jack W. 1995. The Forest Service Program for Forest and Rangeland Resources: A Long-term Strategic Plan. Draft 1995 RPA Program, Forest Service, U.S. Department of Agriculture, Washington, DC. October.

U.S. Department of the Interior. 1991. Report on Our National Parks: Preserving a Priceless Heritage. Research and Forecasts, Inc., for National Park Service, 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., R. Derek Bjonback, Richard A. Aiken, and Donald H. Rosenthal. 1990. "Estimating the Public Benefits of Protecting Forest Quality." Journal of Environmental Management 30:175-189.

Walsh, Richard G., and Lynde O. Gilliam. 1982. "Benefits of Wilderness Expansion with Excess Demand for Indian Peaks." Western Journal of Agricultural Economics 7:1-12.

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., John B. Loomis, and Richard A. Gillman. 1984. "Valuing Option, Existence and Bequest Demands for Wilderness." Land Economics 60:14-29.

Walsh, Richard G., and John R. McKean. 1996. "Option and Anticipatory Values of U.S. Wilderness," in Contingent Valuation of Environmental Preferences. Ian Bateman and Kenneth G. Willis (eds.), Oxford University Press, London. Forthcoming.

Walsh, Richard G., George L. Peterson, and John R. McKean. 1989. "Distribution and Efficiency Effects of Alternative Recreation Funding Methods." Journal of Leisure Research 21:327-347.

Walsh, Richard G., Larry D. Sanders, and John B. Loomis. 1988. "Opportunity Costs of Increased Wilderness Designation." Paper presented at Benchmark 88: A National Outdoor Recreation and Wilderness Forum. Forest Service, U.S. Dept. of Agriculture, Tampa, FL.

Welsh, Michael P., Richard C. Bishop, Marcia L. Phillips, and Robert M. Baumgartner. 1995. Glen Canyon Environmental Studies: Non-use Value Study. Report to the Bureau of Reclamation by Hagler Bailly Consulting, Madison, WI.

Whitehead, John C., and Glenn C. Blomquist. 1991. "A Link Between Behavior, Information, and Existence Values." Leisure Sciences 13:97-109.

A META-ANALYSIS OF RECREATIONAL FISHING

by

Lisa A. Sturtevant, F. Reed Johnson, and William H. Desvousges[1]
Triangle Economic Research
Durham, North Carolina

Abstract

The purpose of this study is to quantitatively synthesize the freshwater recreational fishing demand literature to demonstrate the feasibility of using meta-analysis as a benefits-transfer method.

[1] The authors would like to thank Sarah E. Holden and Theodor P. Tatos for assistance in compiling the data used in this paper and Kristy E. Mathews for help on earlier work associated with this study.

I. INTRODUCTION

Meta-analysis is a process of statistically synthesizing studies of a common subject in order to integrate the findings and draw conclusions. Meta-analysis can be used as a benefits-transfer method to determine benefits associated with the use of environmental resources. Advantages of meta-analysis as a method of benefits transfer include the ability to incorporate results from a range of studies and control for resource-specific characteristics and study assumptions.

Although meta-analysis has been applied in numerous areas of psychology, sociology, and health sciences, it has not played as significant a role in economics. Few economists have used this method of analysis to synthesize results from recreation demand studies. Smith and Kaoru (1990) and Walsh et al. (1992) are notable exceptions. Smith and Kaoru conducted a meta-analysis of travel- cost method (TCM) studies that included a range of recreational activities: water-based recreation (defined as swimming, boating, and fishing), hunting, wilderness hiking, and developed camping. Walsh et al. conducted a meta-analysis of unit-day values for outdoor recreation activities, ranging from fresh and saltwater fishing to winter sports to nonconsumptive wildlife activities. The Walsh et. al meta-analysis included TCM, contingent valuation method (CVM), and hedonic studies.

Both Smith and Kaoru and Walsh et al. demonstrate the feasibility of applying meta-analysis to recreation-demand literature. In each study, the authors were able to draw some general conclusions about what factors are important in determining values for resources providing these recreation services. However, neither of these studies attempts to deal comprehensively with the panel nature of recreation-demand literature. Smith and Kaoru state that the panel nature of the data used in meta-analysis does cause unique problems, but they do not attempt to estimate a panel model. Walsh et al. simply estimate an OLS model without referring to the correlated error structure of the data. In this paper we use panel-estimation techniques to model the data.

Unlike previous meta-analyses, our study focuses on a particular activity and valuation method that eliminates differences in activities and valuation method as a source of variation among studies.

Also, to illustrate the potential uses of meta-analysis, this paper compares the meta-analysis results with those from a site-specific benefits transfer.

II. BENEFITS TRANSFER AND META-ANALYSIS

Benefits-transfer techniques are used when limited time and/or resources prohibit conducting an original study. The collection of original, site-specific data is a time-consuming, expensive task and researchers often need a simpler, more cost-effective way of measuring benefits (or losses) associated with a change in the quality of a resource.

In a benefits-transfer study, estimates from existing studies are used to value resources that have not been studied. Benefits-transfer techniques often have been used to value services provided by natural resources (e.g., Bingham et al., 1992). Researchers often use benefits-transfer methods to value a change in the quality of a resource or to evaluate the benefits of proposed environmental policies.

There are three methods of transferring benefits: basic benefits transfer, benefits-function transfer, and meta-analysis. In the basic benefits transfer, the researcher attempts to select the "best" study from the relevant literature and transfers the estimated value from that study to the current site. The benefits-function transfer adjusts the original study estimates for differences in site characteristics and demographics of users at the current site using model parameters. Meta-analysis uses estimates from multiple studies as observations in a regression to estimate a transfer function.

There are limitations associated with the first two benefits-transfer methods. First, it is extremely difficult— if not impossible— to locate original studies that value a resource identical to the resource of interest. Second, because generally only one study is used in the basic and benefits-function transfer, characteristics of that study will greatly influence the transfer value. If the study methodology or assumptions are flawed or are irrelevant to the benefits-transfer situation, then the transfer value will be biased. Furthermore, original studies generally are not conducted with benefits transfer in mind. Variable definitions and other model details often are not presented in the study.

In a meta-analysis, value estimates are regressed against explanatory variables, which include resource characteristics, study assumptions, and other factors. These factors can be controlled in a site-specific benefits-transfer application by turning "on" and "off" variables in the regression model appropriate to the resource and situation of interest. Thus, only the relevant variables are allowed to affect the estimation and produce a site-specific value estimate. Because meta-analysis relies on multiple studies, analysts can minimize the effect of methodological flaws on value estimates by quantifying the influences of these flaws on the value estimates and using that information appropriately in the benefits transfer. Thus, meta-analysis offers promise as an alternative to *ad hoc* methods of estimating values in many benefits-transfer situations.

III. DEFINITION OF THE PROBLEM

We focused our study on recreational fishing for several reasons. First, recreational fishing is an intensively studied activity. Several recreational fishing bibliographies have been compiled (e.g., Freeman, 1995), which indicates the quantity of literature available on the subject. Second, recreational fishing is an important policy-relevant activity and fishing values are estimated for many situations (e.g., U.S. DOI, 1994). We focus specifically on freshwater fishing, excluding marine fishing because of difficulties in controlling for differences in such factors as species sought and location of the resource.

The demand for recreational fishing resources has been measured using both TCM and CVM models. The TCM is most prevalent, which allowed us to limit our meta-analysis to TCM studies. The structure, theory, and assumptions vary greatly between TCM and CVM studies. By including only TCM studies, we can reduce the variance to be explained. Furthermore, our meta-analysis attempts to explain variation in per-unit values (i.e., per-day or per-trip values), rather than values associated with a change in quality (e.g., increased catch rate). TCM studies are designed to estimate per-day or per-trip values specifically, while CVM surveys often attempt to elicit values for a specified change in a resource.

The TCM emerged as a means of estimating demand for recreational services over 40 years ago. Since then, the TCM has undergone extensive modifications and improvements, and therefore, the sum of the recreational fishing literature includes TCM studies of varying quality and applications. Because TCM methodology has changed significantly over time, we include only studies conducted after 1980. A meta-analysis of the values from these studies can use all of the information that is consistently reported in the studies, simultaneously controlling for the characteristics of the resource and the quality and assumptions of the study.

Nevertheless, creating a meta-analysis data set is difficult because studies seldom report complete information about the survey instrument, data collection, and model estimation. Furthermore, studies report information inconsistently. Our review of more than 100 published and unpublished studies of recreational fishing revealed that only 26 freshwater-fishing TCM studies provided sufficient information to be included in the meta-analysis. (See Table 1 for a listing of the studies used in the meta-analysis). Studies were excluded if there were no value estimates reported, or if insufficient information was provided about the resource or estimation procedure and results. In several cases, no model results were reported, including sample size, coefficient estimates and goodness-of-fit statistics. Clearly, for meta-analysis to reach its full potential in benefits transfer, there needs to be greater standardizing in reporting.

IV. VARIATION IN RECREATIONAL FISHING VALUES

The recreational-fishing values in this meta-analysis are measured in terms of consumer surplus (CS) per unit of fishing. In TCM studies, a demand curve first is calculated using travel and other costs as a proxy for the price of the activity. Then, the area above the price (usually the average price) and below the demand curve is calculated as the CS value. Consumer surplus can be calculated in TCM studies either in per-day or per-trip units. Travel and associated costs will likely be different for a day of fishing versus a fishing trip, if the trip is longer than a day. One-day fishing trips are likely to differ from multiple-day trips because the latter may offer experiences such as camping, sleeping on a boat, or early-morning and late-night fishing that the former would not. Because the per-trip studies generally do not provide information on the average number of days per trip, it was not possible to standardize values on a

Table 1. Studies Used in Meta-Analysis

Authors (date)	Resource	CS (1994 $)
Brown (1983)	Freshwater sportfishing in Lakes Erie and Ontario, and Niagara and St. Lawrence rivers	$1.26 – $67.45
Brown and Shalloof (1986)	Oregon and Washington sportfishing	$31.37 – $84.13
Brown, Sorhus, and Gibbs (1980)	Oregon and Washington sportfishing	$41.09 – $126.75
Brown et al. (1983)	Sportfishing in Rogue River (OR)	$47.27 – $133.66
Donnelly et al. (1984)	Trout fishing in Idaho	$36.30 – $48.31
Duffield, Loomis, and Brooks (1987)	Freshwater sportfishing in Montana	$6.45 – $127.68
Dutta (1984)	Sportfishing in Lake Erie	$2.99 – $10.22
Hushak, Winslow, and Dutta (1988)	Sportfishing in Lake Erie	$0.35 – $8.48
Kealy and Bishop (1986)	Sportfishing in Lake Michigan	$26.91 – $101.24
Loomis, Sorg, and Donnelly (1986)	Sportfishing in Idaho	$47.81 – $91.76
Martin, Bollman, and Gum (1982)	Sportfishing in Lake Mead (CO)	$31.70 – $47.45
Menz and Wilton (1983)	Sportfishing in St. Lawrence River and Chautauqua Lake (NY)	$21.96 – $109.05
Miller and Hay (1984)	Freshwater sportfishing in Idaho, Minnesota, Arizona, Maine, and Tennessee	$7.44 – $99.70
Mitchell and Wade (1991)	Sportfishing in California reservoirs	$22.03 – $37.18
Mullen and Menz (1982)	Freshwater sportfishing in New York	$33.00 – $58.43
Palm and Malvestuto (1983)	Sportfishing in West Point Reservoir (AL/GA)	$11.15 – $55.05
Samples and Bishop (1985)	Sportfishing in Lake Michigan	$0.52 – $30.75
Sorg and Loomis (1986)	Sportfishing in Idaho	$21.95 – $40.49
Vaughan and Russell (1982)	Freshwater fishing in entire U.S.	$14.29 – $39.78
Wade et al. (1989)	Sportfishing in California reservoirs	$27.71 – $44.12
Winslow (1983)	Sportfishing in Lake Erie	$27.22 – $42.08
Ziemer, Musser, and Hill (1980)	Freshwater fishing in Georgia	$41.76 – $161.42

per-day basis.[2] In the analysis, we estimate the differential effect of a per-trip versus per-day measurement by including a shift variable (see Smith and Kaoru, 1990). Such a dummy variable indicates the average effect of per-trip measurement across all observations.

It is reasonable to expect estimated CS values to vary depending on characteristics of the fishing resource. We assume:

$$CS = f(RESOURCE, SPECIES, MODE, LOCATION, USERCHAR, UNIQUE) \quad (1)$$

where

RESOURCE = the body of water (e.g., lake, river, stream);

SPECIES = the species of fish sought (e.g., trout, bass, gamefish, warm water fish);

MODE = the mode of fishing (e.g., boat, shore, pier);

LOCATION = the geographic location of the resource (e.g., southwest, Maine, urban area);

USERCHAR = the characteristics of the users of the resource (e.g., income, education, race); and

UNIQUE = the "pristineness" or "uniqueness" or the resource.

As shown in Smith and Karou (1990), study characteristics also may influence CS estimates. Study characteristics include the structure of the study and assumptions made by the researcher:

$$CS=f(STRUCTURE, ASSUMPTIONS) \quad (2)$$

where in this analysis STRUCTURE reflects the structure of the study and ASSUMPTIONS reflects the researcher's assumptions. STRUCTURE takes into account components of the study, such as travel-cost methodology used, type of survey instrument, and sample size. STRUCTURE also refers to decisions regarding the underlying theory of travel-cost models, such as the treatment of substitutes and the opportunity cost of time. ASSUMPTIONS refers to the researcher's decisions about functional form and estimation technique. In a simple benefits transfer, the effects of these study characteristics are taken

---

[2] We considered using an outside source, such as the *National Survey of Hunting, Fishing, and Wildlife-Associated Recreation* to estimate the average number of days per trip, but we were not confident that this source would be

into account only in a qualitative way. However, in a meta-analysis, we can quantify the effects of these study characteristics.

Explanatory Variables

Table 2 summarizes the explanatory variables included in our analysis. GRLAKE and RIVER are dummy variables for great lakes and river/stream fishing, respectively. We separated Great Lakes fishing from other lake fishing because we want to test the hypothesis that the values associated with fishing on the Great Lakes are significantly different than for other lake fishing. Great Lakes fishing may be different for several reasons, including the sheer size of the lakes relative to other lakes.[3]

Table 2: Explanatory Variables in Meta-Analysis

Variable	Mean	Description
PERTRIP	0.22	Dummy variable = 1 if CS value is per-trip value, 0 if it is a per-day value.
RIVER	0.40	Dummy variable = 1 if river or stream resource is valued, 0 otherwise.
GRLAKE	0.15	Dummy variable = 1 if Great Lake resource is valued, 0 otherwise. (Omitted category is non-Great Lake lakes.)
GAMEFISH	0.19	Dummy variable = 1 if fishing for gamefish (trout, salmon) is being valued, 0 otherwise.
BOTTOM	0.01	Dummy variable = 1 if fishing for bottom fish (carp, catfish) is being valued, 0 otherwise. (Omitted category is other species (walleye, pike) and non-target fishing.)
EAST	0.18	Dummy variable = 1 if resource is in the east, 0 otherwise.
SUBSINC	0.10	Dummy variable = 1 if the price of substitutes is included in the study's demand model, 0 otherwise.
FRACTION	0.26	Fraction of the wage rate used to value the opportunity cost of time.
SEMILOG	0.15	Dummy variable = 1 if semilog functional form, 0 otherwise.
DOUBLE	0.53	Dummy variable = 1 if double log functional form, 0 otherwise. (Omitted category is linear.)
ML	0.04	Dummy variable = 1 if maximum likelihood estimation, 0 otherwise.
OTHER	0.18	Study quality rating based on study characteristics.

representative of each study sample.

[3] The *National Survey of Hunting, Fishing, and Wildlife-Associated Recreation* separates Great Lakes fishing into a separate category of freshwater fishing.

We include two species-specific dummy variables, GAMEFISH and BOTTOM, as explanatory variables. GAMEFISH includes trout and salmon, and BOTTOM is comprised of carp and catfish. Although several other species were mentioned in the studies we reviewed (e.g., pike, walleye, bass), there were only one or two observations associated with each of these species. Our *a priori* expectation is that recreational-fishing experiences where gamefish is the species sought will be valued more highly than carp and catfish fishing. We also expect that targeting a specific species at all, whether gamefish or bottom fish, may result in a higher-valued recreational-fishing experience.

We include a geographic dummy variable, EAST, as an explanatory variable. This geographic variable reflects aspects of resources and anglers in general in that region compared to other regions.[4] We have no *a priori* expectations about the sign on the EAST variable. Other desirable resource characteristics had to be excluded from the model because of insufficient information. These include mode of fishing, characteristics of the users of the resource, average distance traveled by users of the resource, and a measure of the "pristineness" or "uniqueness" of the resource in terms of the quality and quantity of the substitutes available.

We also include six explanatory variables reflecting study characteristics and assumptions. Two are related to the theory underlying travel-cost models, while the others apply to researchers' decisions. The first is a dummy variable, SUBSINC, that takes a value of one if the price of substitutes somehow is incorporated into the demand model. Consistency with economic theory dictates that the price of substitutes should be included in the model. Substitutes typically are limited to other recreational fishing sites, and some travel-cost methods do a better job than others of incorporating substitute prices. For example, random-utility models do the best job of incorporating substitutes, while single-equation travel-cost models at most use *ad hoc* adjustments. The SUBSINC variable simply indicates whether substitutes were incorporated in any way, not the particular way in which they were included.

[4] We have defined the east region as Maryland, New York, Maine, and Delaware.

The treatment of the opportunity cost of time also is important to the underlying theory of travel-cost models. Travel costs should include both monetary costs associated with travel (e.g., gas, tolls) and nonmonetary costs, specifically the opportunity cost of travel time. Some researchers estimate the opportunity cost of time as some fraction of the wage rate, while others do not include it at all. We include FRACTION as an explanatory variable, which is the fraction of the wage rate used in the demand model to value the opportunity cost of time.

In addition to the SUBSINC and FRACTION variables, we include variables for the functional form and model specification chosen by the researcher. We include SEMILOG and DOUBLE dummy variables, which take a value of one for those functional forms. We then can compare the effect of these forms to linear models. Similarly, we include ML and OTHER dummy variables for different estimation techniques. ML takes a value of one for maximum-likelihood procedures for truncated samples. OTHER includes all other estimation techniques.[5]

We hypothesized that the type of TCM used would be important in explaining the variation in CS estimates. However, variables for zonal and individual travel cost are insignificant. This result apparently is caused by having already incorporated important features of these methods as separate explanatory variables. For example, substitutes more often are incorporated into individual travel-cost models than zonal travel-cost models. Zonal travel-cost models more often are estimated using the double log functional form. In addition, we also attempted to include variables for such model features as sample size and survey method. None of these variables was significant in determining CS values. Because of reporting inconsistencies, we were not able to include information about response rates, recall issues, or other biases resulting from data collection techniques. It is possible that such factors could be significant in explaining variation in CS values.

[5] Included in this group are GLS, NLLS, and SUR.

V. MODEL ESTIMATION

One of the primary objectives of this paper is to account for the panel nature of published CS estimates. There are multiple observations from each study, ranging from 2 to 75 observations. Correlation among estimates from the same study violates OLS assumptions. These multiple CS estimates from each study relate to different resources, model specifications, or study assumptions. However, there may be characteristics about each study that systematically affect the value estimates. We report results from three panel models: fixed-effects, random-effects, and separate-variances models.

Panel models require that the data be separated into groups and an index be assigned to each group. In some panel data, these groups are easy to identify. For example, if the data contain profit information from a sample of companies over a certain number of years, then each company is a group. In our case, observations are most obviously grouped by study. However, there may be other similarities that a number of observations share, and it might make sense to more specifically define the groups. Thus, we defined groups by both study and body of water so that the body of water is constant across all observations within a group.[6] Our data are divided into 26 groups with the number of observations ranging from 1 to 42.

OLS Versus Panel Models

A simple OLS regression takes the form:

$$y_t = a + Bx_t + e_t \tag{3}$$

where $t = (1,2,...T)$, a is a constant intercept term, and e is the random error term with $e \sim N(0,\sigma^2)$. OLS treats each observation independently and does not take into account properties inherent to each group. These group effects can possibly influence the estimation in three ways:

- Through individual group intercept terms

[6] We divided the data this way after first defining groups by study alone, but we had difficulties explaining the effect of the body of water because of a few outlier values.

- Through separate group error terms in addition to a common error
- Through separate group error terms with no common error

The fixed-effects model (FEM) assumes that group-specific effects influence the regression through the intercept term:

$$y_{it} = a_i + Bx_{it} + e_t \tag{4}$$

where i = (1,2,...I) for the number of groups. The intercept term varies for observations from different groups, but the error term is common to all observations across all groups.

The random-effects model (REM) assumes that it is not the intercept, but rather part of the random error that varies across studies:

$$y_{it} = a + Bx_{it} + u_i + e_t \tag{5}$$

where u_i is the random error associated with the individual group with u~ N(0, σ_e^2), e is the common error term with e~N(0, σ_e^2), u~N(0, σ_u^2), and $V\sigma_{eu}^2 = 0$. Within-group disturbances are correlated by virtue of their common study component,

$$\text{Corr}[u_i + e_{it}, u_i + e_{is}] = p = \sigma_u^2 / \sigma_e^2 \tag{6}$$

The REM is a two-step procedure. First, the variance components are estimated using the residuals from OLS. Then the REM is estimated using these estimated variances.

The separate-variances model (SVM) assumes a separate random error term associated with each group and no common error term:

$$y_{it} = a + Bx_{it} + e_{it} \tag{7}$$

where e_{it} is the random error associated with each group. The SVM is estimated using maximum-likelihood.

VI. RESULTS

Table 3 presents results from the OLS, FEM, REM, and SVM estimates[7]. Note that several explanatory variables drop out of the fixed-effects model because the intercept term captures within-group variation for these variables. The constant reported in the FEM is the weighted average of the group constants. The FEM estimates shifts in the regressions intercept associated with each group. These shifts are specific to each group and therefore it is not appropriate to use the results to

Table 3: Model Results

	OLS	FEM	REM	SVM
Constant	2.8045*** (0.1905)	3.0701*** (0.2832)	2.8231*** (0.5669)	3.031*** (0.1765)
PERTRIP	-0.0662 (0.1901)		-0.1970 (0.5555)	-0.016 (0.1982)
RIVER	0.3012** (0.1202)		0.0602 (0.5433)	0.2733*** (0.1076)
GRLAKE	-0.0167 (0.2233)		-0.5979 (0.7203)	-0.1362 (0.2502)
GAMEFISH	0.3755*** (0.1445)	1.0000*** (0.2406)	0.8085*** (0.2121)	0.3672*** (0.1323)
BOTTOM	-0.6265 (0.4202)	-0.3017 (0.4331)	-0.3878 (0.4139)	-0.6441** (0.3458)
EAST	-0.3044** (0.1286)	-0.1261 (0.2292)	-0.1895 (0.2019)	-0.3689*** (0.1168)
SUBSINC	-1.6064*** (0.1932)	-0.8652*** (0.2054)	-0.9321*** (0.1974)	-1.6533*** (0.2548)
FRACTION	0.1470 (0.2638)	0.9814*** (0.3735)	0.8101** (0.3432)	-0.0587 (0.2525)
SEMILOG	0.4918** (0.2223)	-0.0545 (0.4292)	0.1100 (0.3607)	0.3403** (0.2038)
DOUBLE	0.6733*** (0.1761)		0.5147 (0.5674)	0.4661*** (0.1740)
ML	-1.1661*** (0.2386)		-0.3101 (0.8636)	-1.2068*** (0.3184)
OTHER	0.0457 (0.1441)		-0.1767 (0.5099)	0.0666 (0.1153)
Lagrange Multiplier chi-squared statistic: 10.947 p-value: 0.0009				

n=258, standard errors in parenthesis
* indicates significance at the 10-percent level
** indicates significance at the 5-percent level
*** indicates significance at the 1-percent level

[7] The SVM is estimated with groups defined by body of water, as opposed to study and body of water.

predict outside of the group. We instead report a weighted average of all groups' intercept terms. The Lagrange Multiplier test rejects the hypothesis that individual error components do not exist.[8]

In each of the models, the PERTRIP dummy variable is insignificant. A possible explanation is that freshwater sportfishing trips predominantly are only one day long. The convention of defining duration as *day* or *trip* may be more a matter of semantics than an actual indication of trip length.

RIVER is significant and positive in the OLS and SVM estimates, but insignificant in the random-effects model. The GRLAKE variable is negative, but insignificant across all models. This might be an indication that, when studying freshwater sportfishing, the body of water is not an important factor in determining values, although a more likely explanation involves possible limitations of the panel models. In the REM, it may be difficult to isolate the effects of variables with no within-group variation, such as RIVER and GRLAKE. This possibility will be discussed later.

GAMEFISH is consistently positive and significant across all models, which agrees with our expectations that more highly prized species will have a positive effect on CS values. Though negative, BOTTOM is only significant in the SVM.

The geographic dummy variable, EAST, is negative and significant in the OLS and SVM estimations, but is insignificant in the other panel models. The negative coefficient may indicate that there are more urban areas in our east region, and therefore resources may not be as pristine as in less urban regions. The EAST specification may be indicative of influences aside from or in addition to geographic factors. A better measure of components of geographic location would be to include specific explanatory variables, such as percentage of the sample area that is urban or sample demographics. However, researchers generally do not report this type of information in their studies and we had no systematic way to infer this information.

[8] R-squared statistics are not reported. They do not provide useful comparisons between OLS and the group-effect models because the group-effect models do not conform to OLS assumptions.

All of the study characteristics variables, with the exception of FRACTION, are significant in the OLS model, while only SUBSINC and FRACTION are significant in the FEM and REM. Some of the functional form and model specifications variables are significant in the SVM. The SUBSINC variable is consistently negative, which correspondents to our expectations that including the price of substitutes in the demand model will result in lower value estimates for that particular fishing resource. FRACTION is positive, indicating a higher opportunity cost of time should be associated with higher value estimates.[9]

The residual plots (Figures 1-3) for the panel models provide information to help us determine which panel estimator is the most appropriate for purposes of predicting values. According to the Loess plot, the FEM (Figure 1.) does not do a good job fitting the data. Because the FEM uses the weighted average intercept term, much of the explanatory power of the regression is lost. The FEM overpredicts at the low end and underpredicts at values in the middle and upper end. The REM (Figure 2.) provides a better fit, while the residuals for the SVM (Figure 3.) are even smaller. The REM underpredicts at extreme values and does a better job, though slightly overpredicts, at values in the middle. The SVM offers the best fit, underprcdicting extremely low values, but fitting the data fairly well at other values.

VII. BENEFITS-TRANSFER COMPARISON

The final objective of this meta-analysis is to provide researchers with an enhanced tool to use in benefits-transfer studies of freshwater recreational fishing. To that end, we apply the results from the three panel models to predict CS values for a day of fishing at three example resources. These examples are designed to contrast results from meta-analysis and basic benefits transfer.

[9] The actual wage rates to which these fractions were applied were not reported in the studies. Having that information would have provided a better estimate of how CS values are affected by the treatment of the opportunity cost of time.

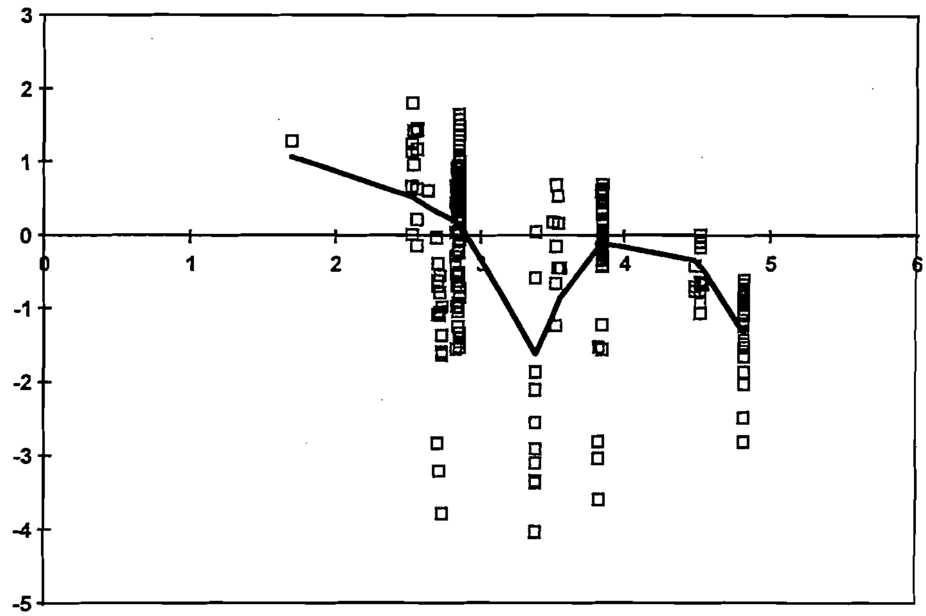


Figure 1. Loess Plot - FEM.

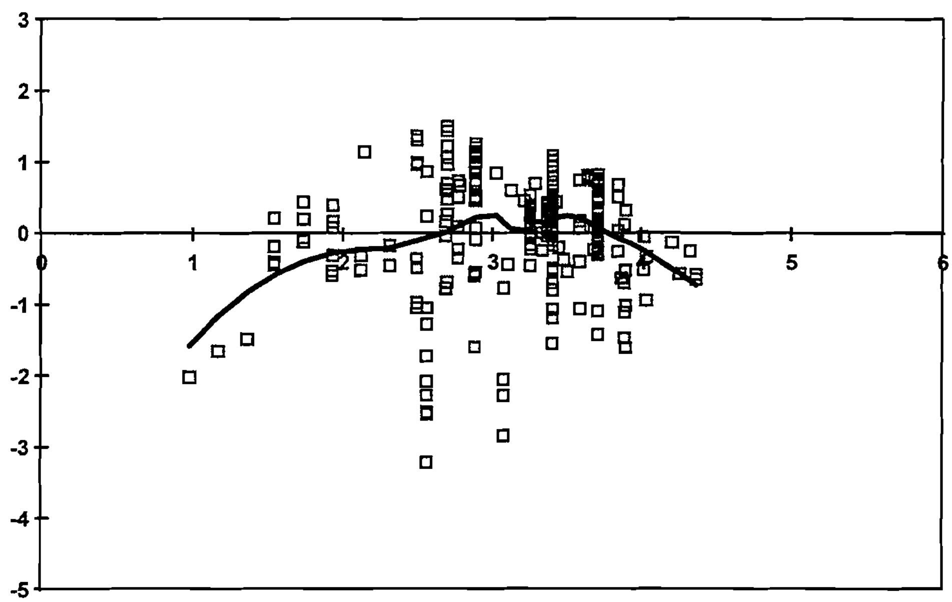


Figure 2. Loess Plot - REM

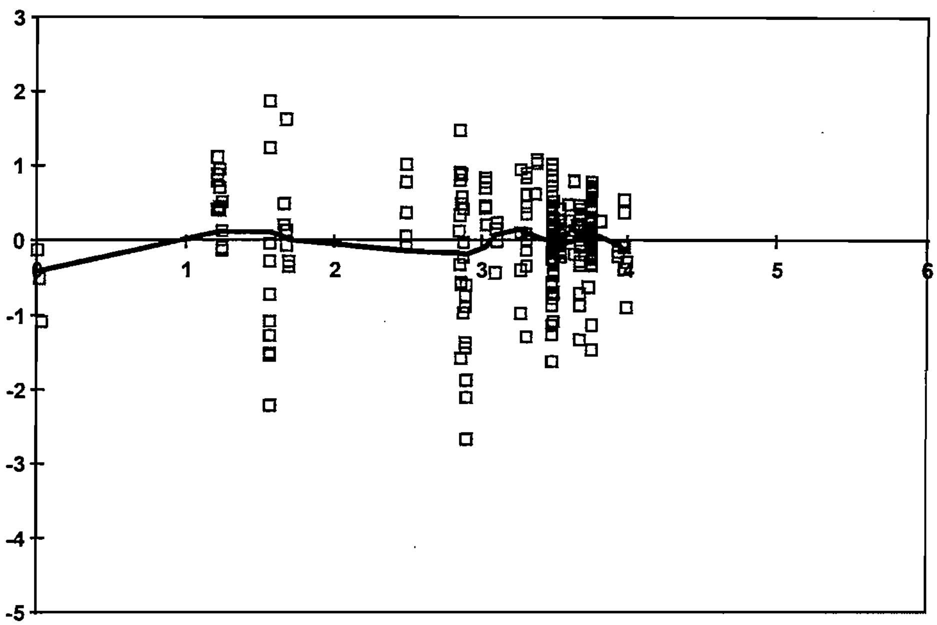


Figure 3. Loess Plot - SVM

We substituted the mean values for explanatory variables for which there is no "best" option (e.g., functional form, model specification, and fraction of the wage rate). In addition, we assign the dummy variable SUBSINC to be one and PERTRIP to be zero in each case. Table 4 summarizes predictions from the FEM, REM, and SVM, along with 90-percent confidence intervals. The final column is a comparison with values that could have been selected by a researcher conducting a basic transfer. The basic values are the mean CS estimates from all studies in our meta-analysis that value a resource similar to each of the three examples.

Table 4: Benefits-Transfer Example - CS/Day Predictions[10]

Resource	FEM[a]	REM[a]	SVM[a]	Basic Transfer[b]
Great Lakes fishing for non-gamefish	$6.96 ($1.74 – 27.93)	$3.90 ($1.58 – 9.65)	$2.33 ($1.07-5.05)	$13.01 ($16.94)
River fishing for gamefish in the east	$22.56 ($8.52 – 59.72)	$20.61 ($12.10 – 35.12)	$6.66 (4.22 – 10.50)	$31.04 ($25.54)
Non-Great Lakes lake fishing for gamefish	$25.59 ($11.47 – 57.09)	$23.46 ($14.32 – 38.42)	$7.33 ($4.80 – 11.18)	$40.74 ($37.86)

[a] 90-percent confidence intervals in parenthesis
[b] Standard deviations in parenthesis

The range of values that might result from the basic transfer are very large. For example, choosing a value from among the Great Lakes, nongamefish fishing studies in our meta-analysis, a researcher could choose a value from between 0.35 (Hushak, Winslow, and Dutta, 1988) and $101.24 (Kealy and Bishop, 1986). Choosing from among eastern river gamefishing studies would result in a value ranging from $9.86 to $67.45 (Brown, 1983). Lastly, choosing from among non-Great Lakes gamefishing studies would result in a value in the range of $11.15 (Palm and Malvestuto, 1983) to 161.42 dollars (Zimer, Musser, and Hill, 1980). Using the predicted values from the REM and SVM estimates, along with the 90 percent confidence intervals, results in more precise value estimates.

VIII. DISCUSSION

In this paper we have shown that we can quantify many factors that are important to explaining the variation in CS values for a unit of recreational fishing. We have identified resource and study characteristics that are significant in influencing these values. These factors should be kept in mind when conducting both benefits-transfer and original studies of freshwater sportfishing. There are additional factors that may be significant to determining CS values, however, because of inconsistency and incompleteness in original studies' reporting, these variables are not included in this analysis. In

[10] All values are in 1994 dollars.

order that we may make use of existing studies, researchers need to report more complete and consistent information in their original studies.

We have explicitly addressed the panel nature of meta-analysis data by estimating fixed-effects, random-effects, and separate-variances models. While these methods are an improvement over OLS, there are certain limitations that still need to be considered. As mentioned previously, the FEM estimates shifts in the regression intercept associated with each group. These shifts are specific to each group and therefore it may not be appropriate to use the results to predict outside of the group. To alleviate this problem, we instead use a weighted average of all groups' intercept terms when calculating predicted values.

The FEM looks for within-group variation among all of the explanatory variables and it drops from the estimation all those that do not vary. Therefore, we cannot isolate the separate effects of these variables on the estimation, which makes this estimator less useful for benefits transfer. Furthermore, the coefficients in the FEM are estimated based on the within-group variation. If there is only one group with variation in a given explanatory variable, then that coefficient will be reflective of only that variation. The estimated relationship between that explanatory variable and the dependent variable may not hold for other groups. If the group containing the variation includes outlier values, then the coefficient estimate may be biased.

In the REM, variables lacking within-group variation are not dropped out of the estimation. However, the estimated group random error terms are estimated based on the within group variation of the explanatory variables. Similar to the FEM, if there is variation in only one study, and that study includes possible outlier values, than the random error and coefficient estimates may be biased.

There are several possibilities to alleviate these limitations, though none of them is satisfactory. One way is to ensure that the explanatory variables have sufficient variation in a sufficient number of groups. This can be accomplished by defining groups more broadly. However, that raises problems related to the amount of total variation needed to be explained by the model and the issue of the appropriate number of observations in each group. It is also important to identify any possible outliers in each of the groups so that these outlier values will not lead to biased coefficient estimates. Often this method can introduce additional problems by forcing the limitation of explanatory variables to those with sufficient variation.

Thus, panel models are theoreticlly a good way to model meta-analysis data. However, future work should include keeping in mind the above-mentioned shortcomings and methods to alleviate them.

The final goal of this meta-analysis was to introduce an improved method of benefits transfer to use to value freshwater recreational fishing resources. We have indicated weaknesses related to other benefits-transfer methods and we have shown how meta-analysis has features that lessen the effects of those weaknesses. We have used meta-analysis to estimate values for several benefits-transfer examples and have shown how these estimates are more precise than the possible estimates in a basic transfer.

Meta-analysis shows promise as an improved benefits-transfer technique. This method can be used in estimating values for other services associated with natural resources, such as beach recreation and boating, provided original studies report sufficient resource and study information. Nevertheless, improved standardization of results reporting is needed before it can reach its full potential.

BIBLIOGRAPHY

Bingham, T.H., E. David, T. Graham-Tomassi, M.J. Kealy, M. LeBlanc, and R. Leeworthy. 1992. *Benefits Transfer: Procedures Problems, and Research Needs.* 1992 Association of Environmental and Resource Economists Workshop. Snowbird, UT (June 3-5).

Brown, T.L. 1983. *A Valuation of New York's Great Lakes Sport Fisheries.* Department of Natural Resources. Cornell University. New York Sea Grant Institute.

Brown, W.G. and F.M. Shaloof. 1986. *Recommended Values for Columbia River Salmon and Steelhead for Current Management Decisions.* Orgeon State University.

Brown, W.G., C. Sorhus, B-L. Chou-Yang, and S.A. Richards. 1983. "Using Individual Observations to Estimate Recreation Demand Functions: A Caution." *American Journal of Agricultural Economics* (February).

Donnelly, D.M., J.B. Loomis, C.F. Sorg, and L.J. Nelson. 1984. *Net Economic Value of Recreational Steelhead Fishing in Idaho.* Unpublished.

Duffield, J., J.B. Loomis, and R. Brooks. 1987. *The Net Economic Value of Fishing in Montana.* Sponsored by the Sport Fish and Wildlife Restoration Acts.

Dutta, N. 1984. *The Value of Recreational Boating and Fishing in the Central Basin of Ohio's Portion of Lake Erie.* Ohio Sea Grant College Program, Technical Bulletin OHSU-TB-018.

Freeman, A.M. 1995. "The Benefits of Water Quality Improvements for Marine Recreation: A Review of the Empirical Evidence." *Marine Resource Economics* 10.

Hushak, L.J., J.M Winslow, and N. Dutta. 1988. "Economic Value of Great Lakes Sportfishing: The Case of Private-Boat Fishing in Ohio's Lake Erie." *Transactions of the American Fisheries Society* 117.

Kealy, M.J. and R.C. Bishop. 1986. "Theoretical and Empirical Specifications Issues in Travel Cost Demand Studies." *American Journal of Agricultural Economics* (August).

Loomis, J.B. 1992. "The Evolution of a More Rigorous Approach to Benefits Transfer: Benefit Function Transfer." *Water Resources Research* 28(3).

Loomis, J.B., C.F. Sorg, and D.M. Donnelly. 1986. "Evaluating Regional Demand Models for Estimating Recreation Use and Economic Benefits: A Case Study." *Water Resources Research* 22(4).

Martin, W.E., F.H. Bollman, and R.L. Gum. 1982. "Economic Value of Lake Mead Fishery." *Fisheries* 7(6).

Menz, F.C. and D.P. Wilton. 1983. "An Economic Study of the Muskellunge Fishery in New York." *New York Fish and Game Journal* 30(1).

Miller, J.R. and M.J. Hay. 1984. "Estimating Substate Values of Fishing and Hunting." *Transactions of the North American Wildlife and Natural Resources Conference* 49.

Mitchell, D.L. and W.W. Wade. 1991. *Recreation Forecasts and Benefit Estimates for California Reservoirs: Recalibrating the California Travel Cost Model.* Draft report to Joint Agency Recreation Committee.

Mullen, J.K. and F.C. Menz. 1982. "The Effect of Acidification Damages on the Economic Value of the Adirondack Fishery to New York Anglers." *American Journal of Agricultural Economics* (February).

Palm, R.C. and S.P Malvestuto. 1983. "Relationships Between Economic Benefit and Sport-Fishing Effort on West Point Reservoir, Alabama-Georgia." *Transactions of the American Fisheries Society* 112.

Samples, K.C. and R.C. Bishop. 1985. "Estimating the Value of Variations in Anglers' Success Rates: An Application to Sportfishing." *Marine Resource Economics* 2(1).

Smith, V.K. and Y. Kaoru. 1990. "Signals or Noise? Explaining the Variation in Recreation Benefit Estimates." *American Journal of Agricultural Economics* (May).

Sorg, C.F. and J.B. Loomis. 1986. "Economic Value of Idaho Sport Fisheries with an Update on Valuation Techniques." *North American Journal of Fisheries Management* 6.

US DOI. 1992. *1991 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation.* US Fish and Wildlife Service. Washington D.C.

US DOI. 1994. *The CERCLA Type A Natural Resource Damages Assessment Model for the Great Lakes Environments (NRDAM/GLE)*, Vol. 1, Technical Documentation. Office of Environmental Policy and Compliance Washington D.C.

Vaughan, W.J. and C.S. Russell. 1982. "Valuing a Fishing Day: An Application of a Systematic Varying Parameter Model." *Land Economics* 58(4).

Wade, W.W., G.M. McCollister, R.J. McCann, and G.M. Jones. 1989. "Recreation Benefits for California Reservoirs: A Multisite Facilities-Augmented Gravity Travel Cost Model." Draft submittal to *Journal of Environmental Economics and Management.*

Walsh, R.G., D.M. Johnson, and J.R. McKean. 1988. *Review of Outdoor Recreation Economic Demand Studies with Nonmarket Benefit Estimates, 1968-1988.* Department of Agricultural and Resource Economics, Colorado State University.

Winslow, J. 1983. *Private-Boat Walleye Angling in the Ohio Waters of Lake Erie: An Economic Evaluation.* Ohio Sea Grant College Program, Technical Bulletin OHSU-TB-008.

Ziemer, R.F., W.N. Musser, and R.C. Hill. 1980. "Recreation Demand Equations: Functional Form and Consumer Surplus." *American Journal of Agricultural Economics* (February).

GROUNDWATER, SURFACE WATER, AND WETLANDS VALUATION FOR BENEFITS TRANSFER: A PROGRESS REPORT

by

Alan Randall and Damitha DeZoysa
The Ohio State University

The overall objective of this study was to perform a comprehensive split-sample contingent valuation (CV) study that would estimate benefits of three environmental services: enhancements to groundwater, surface water and wetland habitat.

Sub-objectives were to:

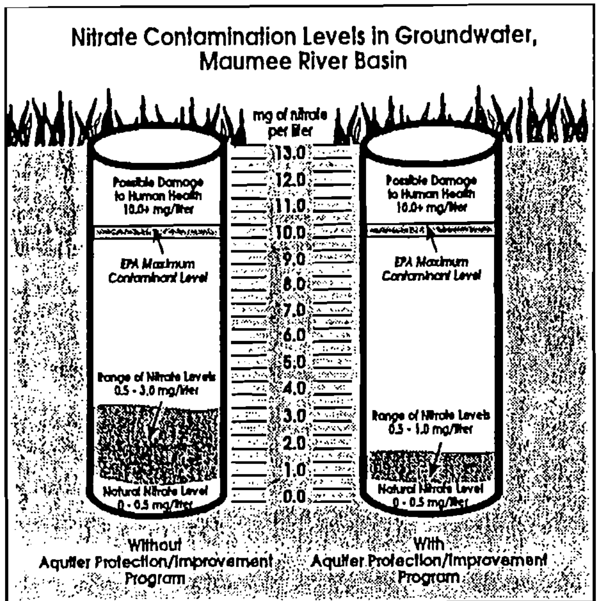
1. Test empirical hypotheses using multivariate analysis of the relationship between vote responses to the offered program and a set of explanatory variables;
2. Test hypotheses concerning the value relationships among components of multipart policies;
3. Develop procedures for benefit transfer when programs are complex; and
4. Contribute to a multistate benefit transfer exercise.

This progress report presents results through subobjectives 1 and 2.

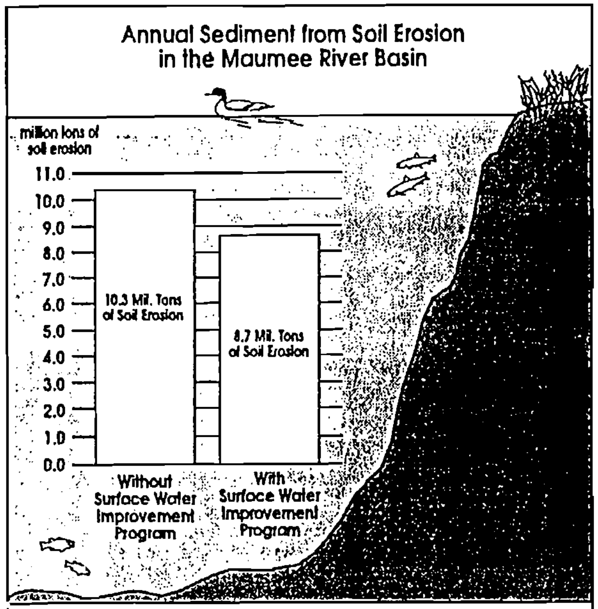
PROGRAMS OFFERED

Three programs were offered: GW, stabilization and reduction of nitrate levels in groundwater in the Maumee River basin in northwestern Ohio; SW, reduction of sediments due to soil erosion, in streams and lakes in the Maumee River basin; and WI, protection and enhancement of wetlands along the shore of the western basin of Lake Erie. The baseline and with-program situations were described in considerable detail in the survey instruments, followed in each case by a box containing a diagram showing baseline and with-program situations and a brief verbal summary (Figure 1).

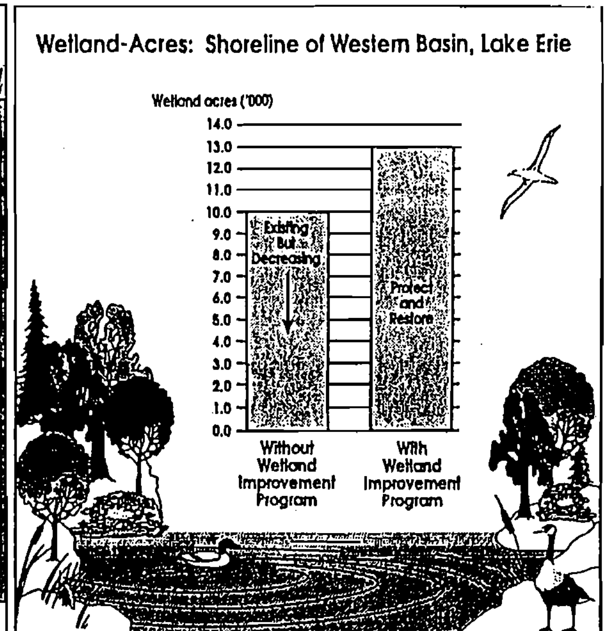
In each case, the program would be financed by a one-time tax with the proceeds dedicated to funding the program. For GW and SW, the funds would provide incentives for farmers to adopt environmentally benign crop-growing practices. For WI, wetlands easements would be purchased.



Without the aquifer protection/improvement program, typical nitrate levels in groundwater are below the EPA 10 mg per liter Maximum Contaminant Level, but are higher than "natural" levels. With the proposed program nitrate levels would be reduced, and eventually stabilized between one-half (0.5) and one mg per liter.



Without the surface water improvement program, the sediment and nutrients going into the Maumee River Basin and Lake Erie would cause deterioration, siltation, water pollution and reduction of fish and wildlife habitat. With the proposed program the volume of sediment entering the Maumee River would be reduced by about 15 percent, thereby reducing the load of sediment in the shipping channel. Recreation in and near the river would be enhanced and the quality of water supplied to Toledo, Maumee and Perrysburg would be improved.



Without the wetland improvement program, wetland acreage and the wildlife habitat would decrease. The proposed program would protect and improve existing wetlands, restore 3,000 additional acres of wetlands, and provide about 20 percent more wildlife habitat for migrating birds and waterfowl.

Figure 1. The programs.

STUDY DESIGN

Three populations were sampled: Maumee drainage rural residents, Maumee drainage urban residents, and residents of Columbus and Cleveland SMAs; the latter sample provided observations of an out-of-region population. Zip-codes in the relevant regions, population-weighted, were selected randomly; then, individuals were randomly selected from Ohio Bureau of Motor Vehicles lists for the selected zipcodes, cleansed of duplicate last names with the same address.. In all, 1050 names were selected; 350 for each population.

The study design was entirely split-sample, with each sample member receiving a single proposal (to provide one, two, or three of GW, SW, and/or WI, as the case may be) and a single tax-price. In all, 147, versions of the survey instrument were used: 3 populations x 7 proposals x 7 prices (Figure 2). Prices used were selected following an open-ended pretest, using the DWEABS method (Cooper 1993) which over-weights prices near the ex ante expected mean and median WTP. Prices used were $0.25, 10, 30, 54, 80, 120, and 200. Ultimately, observations at the $0.25 price were dropped from the analysis; as other researchers have observed, responses at unrealistically low prices tend to be somewhat unstable.

Following a one-shot referendum at a tax-price randomly assigned, open-ended WTP was reported for each respondent. (Only, the referendum results are reported here.) Values obtained are total values, with no formal basis for separating (say) use and passive use values. One may surmise, however, that the out-of-region sample may have been motivated by passive use to a greater degree than the other samples.

SURVEY ADMINISTRATION

Following several focus groups and a field pretest, surveys were mailed to the selected samples. The Dillman total design method was followed, to the extent permitted by a strictly-limited budget. The major items omitted from the Dillman procedure were the use of incentives to respond and the final certified package to persistent-nonrespondents after several mailings and reminder post-cards.

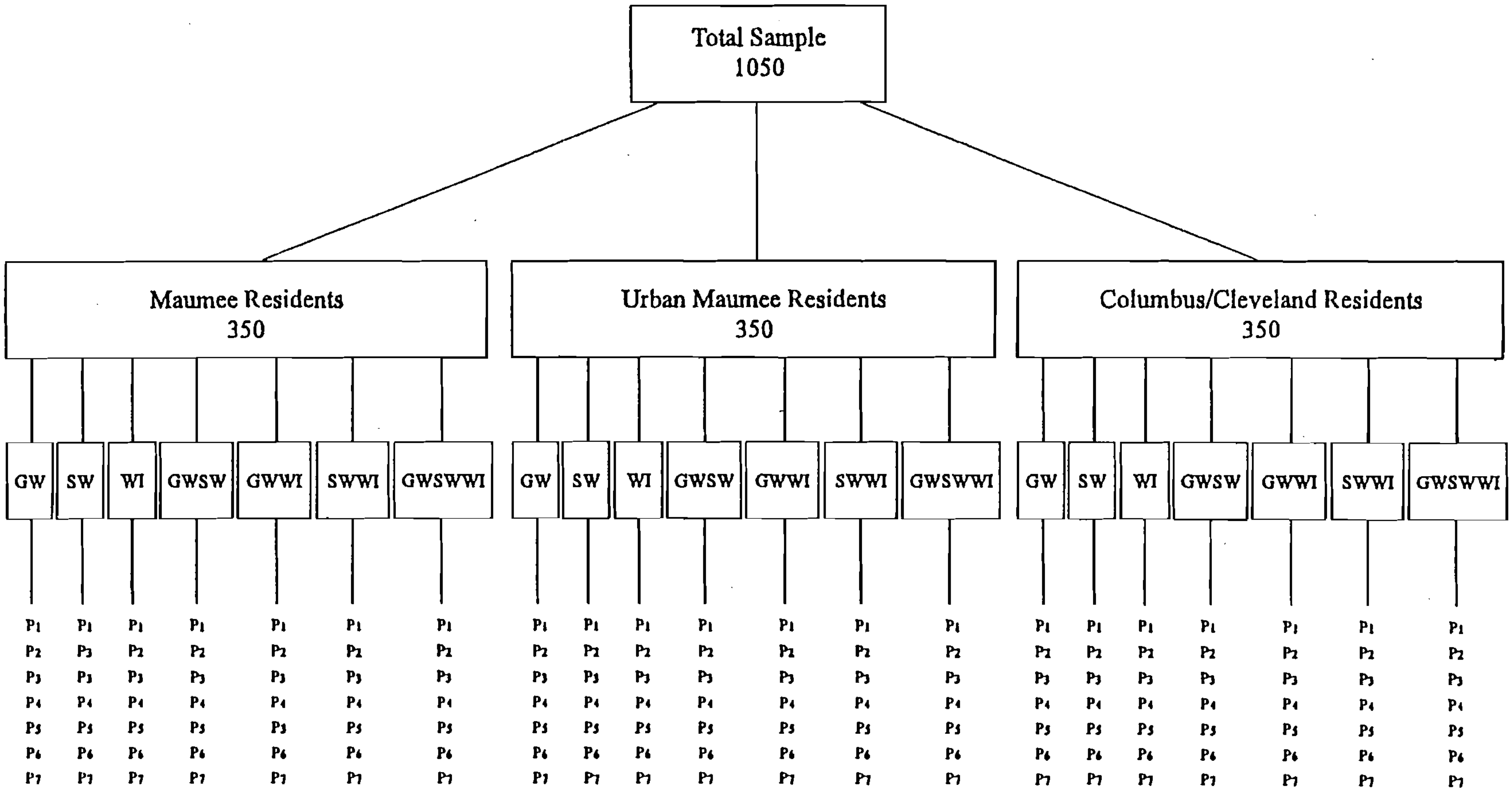


Figure 2. Split-sample study design.

RESULTS

The overall response rate was 51 per cent of delivered questionnaires; for the different populations, response rates were 58 percent for Maumee rural, 50 percent for Maumee urban, and 44 percent for Columbus and Cleveland. After discarding questionnaires with item non-response to the referendum, 427 questionnaires remained: responses to the one-shot referendum were 201 yes, 114 no, and 112 protest-no. This rather large protest-no vote was unexpected. One could guess at the reasons, and three possibilities have some appeal: (1) some respondents were uncomfortable with the idea of subsidizing farmers to use environmentally benign production methods; (2) the multi-county program region was not consistent with the ordinary taxing jurisdictions, i.e., county or state; and (3) in a mail survey, a follow-up question inviting protest-voters to self-identify may actually influence the voting behavior of respondents who "read ahead" before answering.

Standard practice is to discard protest-no voters, treating them as nonrespondents, (i.e. people whose vote does not reflect their WTP). However, our initial econometric analyses of voting behavior suggested that protest-no voters were more similar, attitudinally and demographically, to no voters than to yes voters. This provides a motivation for retaining all no voters including protesters in the sample for subsequent analysis. Accordingly, we provide results for two samples, which we label YN(n = 315, yes: 201, and no: 114) and YNP (n = 427, yes: 201, and no + protest-no: 226). For multivariate analysis, missing data for RHS variables reduced these sample sizes to YN: 286, and YNP: 377.

Multivariate analysis was conducted using the probit choice function and maximum likelihood estimation procedures. (Table 1).

Variables are defined as follows:

LHS	:	VOTE:	yes = 1.
RHS	:	SAMP1:	Maumee rural sample = 1
		SAMP2:	Maumee urban sample = 1
		SAMP3:	Columbus and Cleveland sample = 1
		LPS12 :	log of tax-price interacted with SAMP1 or SAMP2
		LPS3 :	log of tax price interacted with SAMP3
		NI :	household income
		GENDER:	male = 1
		EDUCA1:	not graduate high school = 1
		EDUCA3:	has college degree = 1
		GW1 :	high priority for water quality programs = 1
		GW3 :	low priority for water quality programs = 1
		WHP1 :	high priority of wetlands protection = 1
		WHP3 :	low priority for wetlands protection = 1
		EVH1 :	government should spend more on education, etc. = 1
		EVH3 :	government should spend less on education, etc. = 1
		FU :	expects future visits to program region = 1
		V1 :	GW program = 1
		V2 :	SW program = 1
		V3 :	WI program = 1
		V4 :	GW WI programs = 1
		V5 :	GW SW programs = 1
		V6 :	SW WI programs = 1

Table 1. Multivariate probit analysis.

	YN data			YNP data		
Variable	Parameter Estimate	p-value	Variable Mean	Parameter Estimate	p-value	Variable Mean
CONSTANT	1.0073	0.12		0.50711	0.36	
LPS12	-0.34909	0.00***	2.7985	-0.32652	0.00***	2.9144
LPS3	-0.75556	0.00***	1.0897	-0.71190	0.00***	1.0165
NI	0.000009	0.01***	40808	0.000009	0.00***	39749
GENDER	-0.27629	0.12	0.48601	-0.23578	0.11	0.49867
EDUCA1	-0.04612	0.86	0.13287	0.053960	0.81	0.12997
EDUCA3	0.15117	0.46	0.31818	0.082365	0.63	0.30769
GW1	0.40853	0.02***	0.61538	0.43086	0.00***	0.56499
GW3	-0.32400	0.53	0.034965	-0.49531	0.24	0.055703
WHP1	0.52905	0.01***	0.35664	0.48252	0.00***	0.31300
WHP3	-0.84810	0.00***	0.16434	-0.86892	0.00***	0.21485
EHV1	0.38836	0.03***	0.61888	0.31114	0.04***	0.60477
EHV3	0.037638	0.92	0.052448	0.10803	0.75	0.061008
FU	0.30428	0.12	0.70979	0.36034	0.03***	0.68435
V1	-0.49933	0.13	0.14685	-0.55390	0.04***	0.16446
V2	-0.09856	0.78	0.11888	-0.17550	0.54	0.12732
V3	-0.48867	0.15	0.13986	-0.40309	0.16	0.12997
V4	-0.11509	0.73	0.13986	-0.25431	0.36	0.14589
V5	-0.11694	0.71	0.16084	-0.056697	0.83	0.15650
V6	-0.54204	0.10*	0.15035	-0.34209	0.23	0.14058
SAMP1	0.076831	0.71	0.36364	0.022203	0.89	0.40053
SAMP3	1.6300	0.14	0.28671	1.6777	0.05**	0.26525

Results obtained were consistent with prior expectations with a few exceptions related to significance of the estimates. The price variable (logged, and combined with sample dummies to permit the price response to differ across samples) is the most significant predictor of the respondents' voting behavior. The out-of-region sample exhibited the most price-sensitive voting response. The income coefficient was positive and significant. Attitudinal variables indicating high priority for water quality and wetland protection program, desire to increase public spending on education, health and vocational training programs, and the expectation of future visits to the region were positively correlated with VOTE and statistically significant. All the program dummy variables had the expected sign compared to the omitted program (GWSWWI), confirming monotonicity; i.e., *ceteris paribus* tax-price, more public goods are preferred to less. With YN data, 74 per cent of vote responses were predicted correctly, while with YNP data the corresponding result was 71 per cent (Table 2). However, the YNP model predicted the actual counts of yes vs. no votes more accurately.

Table 2. Predicted vs. Actual votes

		YN Data			YNP Data		
		Predicted		Total	Predicted		Total
		No	Yes		No	Yes	
Actual	No	52	52	104	138	57	195
	Yes	23	159	182	53	129	182
Total		75	211	286	191	186	377

MEDIAN AND MEAN WTP

The standard measures of central tendency for WTP are the median and mean, which, can be interpreted respectively in terms of a voting criterion (WTP of the median voting household) and the potential Pareto-improvement criterion (benefit equals mean WTP aggregated across households). Median WTP/household, estimated with the YNP data set, is reported (Table 3) for each of the seven programs/combinations, pooled across the three samples, and for each of the three samples pooled across programs. As other researchers have observed (e.g. Haab and McConnell 1995), the log-normal probit model often provides a good fit of the vote data within the range of tax-prices assigned, but generates absurdly high estimates of mean WTP. Lower bound means (LBM), which assume effectively that all incremental yes votes as the tax-price is reduced apply only to the lower end-point in each given range of tax-prices, are reported (Table 3).

Table 3. Estimated median and lower bound mean (LBM) WTP ($/household, one-time payment), YNP data.

Program	Sample	Median	LBM
GW	*	20.80	52.78
SW	*	50.27	78.38
WI	*	29.56	62.57
GWWI	*	41.83	72.65
GWSW	*	66.32	87.09
SWWI	*	34.08	66.63
GWSWWI	*	75.70	91.41
**	1	35.27	74.56
**	2	32.96	72.96
**	3	52.45	68.37

* For all samples pooled.
** All program responses pooled.

Observe that while sample 3 (Columbus and Cleveland) had the highest estimated median WTP, it had the lowest LBM. This is consistent with the estimated probit model (Table 2), in which sample 3 has the highest intercept and the steepest tax-price slope.

MULTI-COMPONENT PROGRAMS

The literature on valuation of multi-component programs is replete with empirical reports that WTP for a multi-component program is less than the sum of WTP for its components evaluated independently. What is controversial is the interpretation of this phenomenon. Diamond (1996) and Kahneman and Knetsch (1992), seeing no theoretical reason for this phenomenon, regard it a evidence of a pathology in contingent valuation. However, Randall and Hoehn (1996), Hoehn and Loomis (1993), Hoehn (1991), and Hoehn and Randall (1989) provide arguments that standard economic concepts -- constrained budgets, and substitution among policy components -- can be expected to induce the observed relationship. Randall and Hoehn demonstrate the predicted relationships with numerical simulations using an estimated system of market demands.

Comparing estimated median WTP across programs (Table 3) we observe that GWSWWI is higher-valued than any single-component or two-component program; GWWI is higher-valued than GW or WI; GWSW is higher-valued than GW or SW; and SWWI is higher-valued than WI. However, in every case the value of the multi-component program is less than the sum of the values of its components evaluated separately. The one exception to this pattern of results is SWWI which is lower-valued than SW. This result seems to be an artifact of our small sample size for such an elaborate split-sample design: non-response rate was abnormally high for SWWI at prices $80 and higher (76% vs. 49% for the overall survey), an event which may well have been random.

Using the Wald test, we rejected the following null hypotheses: GW $\geq$ GWSW, and GW = SW = WI = GWWI = GWSW = SWWI = GWSWWI. For SW $\geq$ SWWI, the sign was "wrong," in that the results supported the null rather than the expected (alternative) hypothesis. For all other pairs of single and/or multi-component programs, the signs were correct but the differences were not significant.

VALIDITY OF RESULTS

As is becoming all too common in research at academic institutions, this study was performed with a substantial investment of research effort complemented by a ludicrously small budget for out-of-pocket expenses. The result was an elaborate split-sample research design supported by a relatively thin data set. The predictable outcome was that (until we can afford to collect more data) much of the statistical testing that would be *a priori* desirable has been hampered or precluded in practice by inadequate sample size.

Nevertheless, there are some indications of validity. First, sound CV research practices have been followed. In particular: the split-sample design recommended by the NOAA panel (Arrow et al 1993) was used; the questionnaires were carefully developed and pre-tested; and the programs offered were not merely plausible, in fact baseline conditions were realistic descriptions of actual conditions and with-program conditions were developed with the advice of knowledgeable policy-makers and scientists. Second, the multivariate probit analyses (Table 1) demonstrate construct validity. Third, the relationships among WTP for single and multi-component programs were mostly consistent with theoretical expectations and significant in some cases; this is a fairly strong result, given the split-sample design and the thin data set.

IMPLICATIONS FOR ENVIRONMENTAL POLICY

Estimated median WTP provides a measure of the WTP of the median voting household, whereas LBM provides a lower-bound measure of mean benefits/household. The study generated estimates from two kinds of data sets, YN and YNP, and for three samples representing respectively the Maumee rural (MR), the Maumee urban (MU), and the Columbus-Cleveland (CC) populations. Obviously, some decisions about appropriate aggregation strategies must precede policy pronouncements.

First, we lean toward the LBM as a lower-bound estimate of mean WTP/household, the proper benefit measure for a potential Pareto-improvement test. Second, we lean toward the YNP data set for generating household-level WTP estimates, because there seemed insufficient reason to discard the rather large group of respondents who reported protest-no votes. So, benefit estimates reported below are one-time payments based on LBMs for YNP data, with zero WTP assigned to non-respondent households. Aggregation strategies require more subtle consideration; accordingly, we report results for more than one level of aggregation, and offer some commentary concerning aggregation levels.

Groundwater Program

Aggregating across the in-region population, MR + MU, the benefits of the GW program amount to $4.04 per acre of cropland. Adding-in the CC population generates a benefit estimate of $17.55 per acre. If one was willing to assume that WTP for the CC sample was representative of all households in non-Maumee-basin Ohio, an aggregate benefit of $71.02 would be obtained. Aggregating across just MR + MU populations ignores the strong evidence of positive out-of-region WTP. However, one wonders if out-of-region WTP for the GW program in the Maumee basin would have been so large if respondents were offered simultaneously GW programs in other cropland regions of the state.

Surface Water Program

Benefits per acre of cropland were $6.05 for MR + MU population, $26.06 for MR + MU + CC, and $101.30 for the population of Ohio. Given that sediments in the Maumee River eventually contaminate the western basin of Lake Erie, a popular resort area which draws visitors from much of Ohio and beyond, the broader aggregates have somewhat more credibility in the case of the SW program.

Wetland Improvement Program

Benefits per acre of wetland to be protected amount to $1,077 for MR + MU population, $21,566 for MR + MU + CC, and $85,215 for the population of Ohio. These numbers may seem large, but the wetlands along the shore of the western basin of Lake Erie are a major resource that has already been much diminished, in close proximity to a popular resort area. We feel comfortable assuming a substantial clientele of active and passive users of this resource.

FUTURE RESEARCH

Tasks remaining include further analysis along the lines reported here, analysis of the open-ended WTP responses, in-depth analysis of protest-no responses and nonresponse, and estimation of substitution relationship among the three environmental enhancements offered. Ultimately, this research will contribute to multistate benefit transfer exercises underway in regional research project W-133.

REFERENCES

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

Cooper, J. C. 1993. "Optimal Bid Selection for Dichotomous choice Contingent valuation Surveys." Journal of Environmental Economics and Management, Vol. 24, pp 25-40.

Diamond, P. 1996. "Testing the Internal Consistency of Contingent Valuation." Journal of Environmental Economics and management 30:337-347.

Habb, T. C. and K. E. McConnell 1995. "Referendum Models and Negative Willingness to Pay: Alternative Solutions." Unpublished Paper, Department of Agricultural and Resource Economics, University of Maryland, College Park.

Hoehn, J. P., 1991. "Valuing the Multidiimensional Impacts of Environmental Policy: Theory and Methods." American Journal of Agricultural Economics, Vol. 73, pp. 289-99.

Hoehn, J. P. and A. Randall. 1989 "Too Many Proposals Pass the Benefit Cost Test." American Economic Review, Vol. 79, pp. 543-51.

Hoehn, J. P. and J. B. Loomis. 1993. "Substitution Effects in the Valuation of Multiple Environmental Programs." Journal of Environmental Economics and Management, Vol. 25, pp. 56-75.

Kahneman, D. and J. L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." Journal of Environmental Economics and Management, Vol. 22, pp. 57-70.

Randall, A. and J. P. Hoehn. 1996. "Embedding in Market Demand Systems." Journal of Environmental Economics and Management 30:369-380.

MEASURING THE DIFFERENCES IN MEAN WILLINGNESS TO PAY WHEN DICHOTOMOUS CHOICE CONTINGENT VALUATION RESPONSES ARE NOT INDEPENDENT

by

Gregory L. Poe, Michael P. Welsh, and Patricia A. Champ*

ABSTRACT

Dichotomous choice contingent valuation surveys frequently elicit multiple values in a single questionnaire. If individual responses are correlated across scenarios, the standard approach of estimating willingness to pay (WTP) functions independently for each scenario may result in biased estimates of the significance of the difference in mean WTP values. This paper applies an alternative bivariate probit approach that explicitly accounts for correlation across errors in the estimation of WTP and mean WTP distributions. This correlation is found to have an effect on the significance level of tests of the difference in mean WTP between scenarios mean WTP difference tests.

The authors are indebted to Bill Provencher and Jean-Paul Chavas for raising the independence question in the context of another paper. Richard Bishop provided helpful comments and was instrumental in the funding, design, and implementation of the surveys used in this study. Tim Mount, Carlos Reberte and Joseph Cooper provided programming advice. Participants in the 1996 W-133 meetings and two anonymous reviewers provided useful comments. Of course, none of the above are responsible for any errors in this manuscript. Funding for this project was provided by the College of Agriculture and Life Sciences, Cornell University; College of Agricultural and Life Sciences, University of Wisconsin-Madison; and the Glen Canyon Environmental Studies Project, U.S. Bureau of Reclamation. Cornell ARME Working Paper WP 95-02.

* The authors' affiliations are Department of Agricultural, Resource, and Managerial Economics, Cornell University, Ithaca, NY; Hagler Bailly Consultants, Inc., Madison, WI; and the USFS Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

1. INTRODUCTION

The high costs associated with collecting primary data, coupled with a need for within sample comparisons of Hicksian surplus values, frequently leads researchers to include several contingent valuation (CV) questions in a single survey. In particular, resource valuation surveys often elicit surplus values for a baseline level of resource provision (e.g., current hunting conditions) and then ask about values for alternative levels of provision (e.g,. improved hunting conditions). Although this approach reduces data collection costs and allows for the estimation of continuous resource valuation functions [e.g., Boyle, Welsh, and Bishop, 1993], possible correlation between responses complicates policy relevant comparisons of expected benefits across scenarios. Such complications similarly arise in testing for within subject embedding effects, in which values placed on a comprehensive good are compared with values for a subset of the comprehensive good [e.g., Carson and Mitchell]. Correlation across valuation response functions, or more formally across errors, will be associated with the extent that estimated models fail to capture individual specific factors that have a common effect on responses across questions. Although it is recognized that individual valuation processes may be complex and heterogeneous, most estimated valuation functions consist of relatively simple models in which the combined effects of excluded variables are assumed to be summarized by a random disturbance. If error terms consist of systematic components, unmeasurable or omitted variables that represent factors particular to individuals are likely to create correlation in estimated errors across equations [Hsiao]. The direction of correlation should be affected by the perceived likeness or substitutability of the goods being valued: for goods that are viewed as substitutes, a negative correlation is expected; closely related, embedded, or nested goods would likely result in a positive correlation in error terms. Correlated responses could also be caused by systematic response patterns associated with CV such as "yea saying" [Kanninen], "symbolic" effects [Boyle, Welsh, and Bishop, 1991], "warm glow" and "embedding" effects [Kahneman and Knetsch], or "starting point" biases [Cameron and Quiggin]. In all, many factors are relegated to the error term in analyses of CV responses, increasing the likelihood that errors will be correlated across valuation response categories.

To the extent that individual responses to successive scenarios within a survey are correlated, the standard CV approach of estimating independent willingness to pay (WTP) functions will provide biased estimates of the variance of mean WTP. In turn, the significance of the difference of mean WTP values between scenarios will be biased. Two factors underlying this bias are most easily distinguished by referring to the well known formula for the variance of the difference of two normal distributions: $var(X-Y)=var(X) + var(Y) - 2*cov(X,Y)$. The first factor affecting this difference is an "efficiency effect" associated with the estimation of individual distributions. Although point estimates from independently estimated WTP models remain consistent even if responses are correlated across scenarios, the estimates are inefficient. Because mean WTP distributions, depicted here as X and Y, are derived from coefficients of the estimated WTP functions, the dispersion of individual mean WTP distributions will be biased if the analyst does not account for correlation. There may also be efficiency gains associated with imposing restrictions across equations.[2] The second factor affecting the distribution of the difference is a "correlation effect" in that the failure to account for correlation between X and Y, depicted above by the covariance, will lead to biased estimates of the variance of the difference between these two variables.

Using data from three dichotomous choice CV resource studies as examples, this paper investigates the impact that the efficiency effect has on the estimates of individual WTP distributions, and the combined impact of both the efficiency and the correlation effects on the variance and the significance of the difference of the mean WTP distributions. The remainder of the paper is organized as follows. Section 2 provides the conceptual framework for the bivariate probit and bootstrapping approaches used in the analysis. The CV studies used to investigate these issues are described in Section 3. Empirical results are discussed in Section 4, and Section 5 provides the conclusions from this study and implications for future research.

2. CONCEPTUAL FRAMEWORK

Assume that the ith individual has some true surplus value (s_{ji}) for the good described in the jth scenario, and that the respondent will indicate (I_{ji}=1) that they are willing to pay the posted price (p_{ji}) if $s_{ji} \geq p_{ji}$. If $s_{ji} < p_{ji}$ the individual will not be willing to pay p_{ji}, and $I_{ji}=0$. Following the random utility framework presented in Hanemann (1984) and Cameron and co-authors (Cameron and James; Cameron and Quiggin), assume that the unobserved value $s_{ji} = \beta_j' x_{ji} + u_{ji}$, where the systematic component, $\beta_j' x_{ji}$, is a function of a vector, x_j, of observable attributes of the respondent, including the dichotomous choice posted price, and u_j is an unobservable random disturbance assumed to be distributed $N(0, \sigma_j^2)$.

Standard approaches to evaluating and testing alternative scenarios assume that the u_{ji} are uncorrelated across scenarios[3]. Under this assumption WTP distributions are estimated independently for each scenario. Approximate distributions of mean WTP are derived from these estimated functions by applying bootstrapping or other repeated sampling techniques [Park, Loomis, and Creel]. Under the independence assumption, the difference in approximate distributions of mean WTP can be estimated using an empirical convolutions approach or by directly bootstrapping the difference [Poe, Severance-Lossin, and Welsh].[4]

As suggested previously, the assumption of independence may not be appropriate if multiple CV questions are posed in the same questionnaire. Econometrically, in a manner analogous to seemingly unrelated regressions, this non-independence between the two valuation functions for scenarios one (s1) and two (s2) may be accommodated by explicitly accounting for cross equation correlation in the estimation process. Within a discrete choice format, this can be accomplished by assuming a bivariate normal distribution $BVN(\beta 1'x_1, \beta_2'x_2, \sigma_1^2, \sigma_2^2, \rho)$ of the errors where β_j and x_j correspond to parameters previously defined and ρ is the correlation coefficient. Defining $z_1 = -\beta_1'x_1/\sigma_1$ and $z_2 = -\beta_2'x_2/\sigma_2$ to be standardized normal errors, the standard bivariate normal distribution SBVN(ρ) for (Z_1, Z_2) takes the following form.

$$\Phi(z_1, z_2; \rho) = \frac{\exp^{-(z_1^2 + z_2^2 - 2\rho z_1 z_2)/2(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}} \tag{1}$$

If ρ is indeed zero, this density function collapses to the product of two independent normal density functions, and the univariate approach outlined previously is appropriate for estimating separate probit WTP distributions, independent mean WTP distributions, and the difference of the mean WTP distributions. If $\rho \neq$ o, the associated likelihood function for the four possible pairs of responses (no (s1) - no (s2), no (s1) - yes (s2), yes (s1) - no (s2), yes (s1) - yes (s2)) across equations is given as:

$$\begin{aligned} L = \prod_i \; & \Big[\int_{-\infty}^{\frac{-\beta_1' x_1}{\sigma_1}} \int_{-\infty}^{\frac{-\beta_2' x_2}{\sigma_2}} \Phi(z_1, z_2; \rho) dz_1 dz_2 \Big]^{(1-I_1)(1-I_2)} * \\ & \Big[\int_{-\infty}^{\frac{-\beta_1' x_1}{\sigma_1}} \int_{\frac{-\beta_2' x_2}{\sigma_2}}^{\infty} \Phi(z_1, z_2; \rho) dz_1 dz_2 \Big]^{(1-I_1)I_2} * \\ & \Big[\int_{\frac{-\beta_1' x_1}{\sigma_1}}^{\infty} \int_{-\infty}^{\frac{-\beta_2' x_2}{\sigma_2}} \Phi(z_1, z_2; \rho) dz_1 dz_2 \Big]^{I_1(1-I_2)} * \\ & \Big[\int_{\frac{-\beta_1' x_1}{\sigma_1}}^{\infty} \int_{\frac{-\beta_2' x_2}{\sigma_2}}^{\infty} \Phi(z_1, z_2; \rho) dz_1 dz_2 \Big]^{I_1 I_2} \end{aligned} \tag{2}$$

The hypothesis H_o: $\rho = 0$ can be evaluated with a standard likelihood ratio test, $-2(LL_1 + LL_2 - LL_j) \sim \chi^2$, by comparing the log of this likelihood function (LL_j) with the sum of the log likelihoods (LL_1, LL_2) associated with the independently estimated probit distributions [Greene]. Similarly, a comparison of log likelihood values can be used to assess the validity of cross equation restrictions on the estimated parameters. In making such comparisons, it should be noted that the greatest efficiency gains are expected when $X_1 \neq X_2$. But, in contrast to continuous dependent variables, there should also be efficiency gains even when the covariates are identical across equations [Alberini and Kanninen].

If the null hypothesis H_o: $\rho = 0$ is rejected, $E(\overline{WTP}_2 | \overline{WTP}_1)$ is a non-zero function of ρ [Goldberger] and the mean WTP distributions will depend upon the joint distribution of estimated parameters, one of which is ρ. Consequently, simulated mean WTP distribution values from the joint distribution must be paired, and the difference of the mean WTP distributions $\overline{WTP}_{jB}$ can be estimated by directly bootstrapping the difference,

$$D = \overline{WTP_{1m}} - \overline{WTP_{2m}} \qquad m = 1, \ldots, B. \tag{3}$$

where B is the number of paired bootstrap observations. Following the percentile approach in Efron and Tibshirani, the approximate one-sided significance of the difference is obtained by computing the proportion of negative values in D.

3. DATA

The data for this analysis were taken from three separate dichotomous choice CV mail surveys of recreational resource use. Examples of individual CV questions from each of the surveys are provided in the appendix.

The *Escanaba Lake Survey* was conducted as part of a study to assess the validity of CV values by comparing hypothetical WTP to actual WTP.[5] Escanaba Lake is one of five lakes managed by the

Wisconsin Department of Natural Resources in the Northern Highland State Forest of Vilas County. It is the only lake in Northern Wisconsin where anglers can fish for walleye after the ice is off the lake before the regular fishing season. This early season between "ice-off" and the regular fishing season can vary from a few days to a few weeks. Individuals who had fished the early season at Escanaba Lake in 1989, 1990, or 1991 were mailed a CV questionnaire in March of 1992 (prior to the early season). Eight hundred and twenty questionnaires were mailed and 621 were completed. Adjusting for undeliverable questionnaires, the response rate was 82 percent. The questionnaire included two dichotomous choice CV questions. The first question asked whether the individual would pay $X for a "baseline permit" to fish the upcoming early season, in which expected catch corresponds to historical levels. The second CV question asked whether the respondent would pay $Y for a permit to fish the upcoming early season, if there would be "15 percent fewer" walleye than usual in Escanaba Lake. The format of these questions is that the second CV question ("15 percent fewer") is nested in the baseline case, in that, with exception of the number of fish available, the scenarios are identical.

The 1991 and 1992 *Sandhill Public Deer Hunt Surveys* were part of a larger study to assess the ability of recreationists to recall expenses related to a special deer hunt (see Champ and Bishop for further details). Sandhill Wildlife Demonstration Area is a wildlife research property managed by the Wisconsin Department of Natural Resources in Wood County, Wisconsin. In 1991, 352 one-day deer hunting permits were issued for an either sex deer hunt to be held in November. One hundred seventy-seven of the permit holders were sent questionnaires after the hunt. Seventy of the permit holders who were sent a questionnaire did not attend the hunt at Sandhill. Of the 107 hunters who received a questionnaire and hunted, 104 (97 percent) returned the questionnaire. In 1992, the November Sandhill hunt was for antlerless deer only. Two hundred thirty permits were issued and 117 hunters were sent a questionnaire. One hundred seven (91 percent of the deliverable questionnaires) questionnaires were returned. The questionnaires sent in 1991 and 1992 were very similar. Respondents were asked about their expenses related to the Sandhill deer hunt, the quality of the hunt, some demographic questions, and two dichotomous choice CV questions. One CV question asked about their willingness to pay for an "either sex" deer hunting permit and the other asked about their willingness to pay for an "antlerless" deer hunting permit at Sandhill. As with the *Escanaba Lake Survey*, the "antlerless" deer permit is formally a nested subset of the "either sex" permit. However, since each permit is good for only one kill, an element of substitution arises. Hunters with "either sex" permits typically report that they do not want to "waste" their permit on does and immature animals, and therefore these permits may be viewed as having non-inclusive elements. *A priori* this element of substitutability would be expected to have a negative impact on the correlation coefficient.

The objective of the *Grand Canyon White Water Boater Survey* was to estimate a statistical relationship between Hicksian surplus values for white-water trips and average daily Colorado River flows between 5,000 and 40,000 cubic feet per second (see Boyle, Welsh and Bishop, 1993 for further details). In this survey individual respondents were each asked four dichotomous choice CV questions corresponding to the following hypothetical flow levels: 5, 13, 22, and 40 thousand cubic feet per second (kcfs). Prior to answering the valuation questions, respondents answered a series of questions about the attributes of their Grand Canyon white-water trip, including trip expenditures. Each of the valuation questions were preceded by a description of the boating and camping conditions associated with that specific flow. Conducted in 1986, 169 usable responses were obtained from private boaters, representing approximately 91 percent of deliverable surveys. In contrast to the fishing and hunting surveys, each flow level is associated with distinct characteristics, and one flow level cannot be viewed as a nested subset of other flow levels. However, some attributes associated with different flow levels are common even in paired scenarios that describe substantially different flows. For example, both 5 kcfs and 40 kcfs entail inconvenient portaging around additional rapids. Similarly, adjacent flow levels have trip attributes that overlap considerably, but maintain some distinct elements. To the extent that individuals have preferences over flow characteristics, some flow levels should be regarded as substitutes.

4. RESULTS

Estimated CV responses functions and associated mean WTP values in each of the three surveys were compared with values obtained from different scenarios in the same questionnaire. The procedure for evaluating the effects of cross scenario correlation was to evaluate each pair of questions as follows. First, bivariate (joint) and univariate (independent) probit models were estimated using maximum likelihood techniques. Likelihood ratio tests were used to evaluate the hypothesis that H_0^1: $\rho = 0$ as well as to test various cross-equation equality restrictions. For comparisons in which H_0^1 is not rejected, no additional analyses were conducted beyond the initial maximum likelihood estimates. In the cases where H_o^1 is rejected, 10,000 simulated values of mean WTP were estimated using numerical integration techniques over the non-negative range of the WTP distributions [Hanemann 1984, 1989] and a parametric bootstrap technique that draws simulated coefficient values from the covariance matrix [Park, Loomis and Creel; Krinsky and Robb]. For the jointly estimated bivariate model, the mean WTP distributions for each question were approximated after accounting for ρ in the estimated covariance matrix. In both the joint and independent models, pairwise differences were calculated as in equation (3). Comparisons of these approximate distributions of the difference for the joint and independent estimates provide the basis for assessing the effects of the independence assumption on the distribution of the difference. The approximate one-sided significance of the difference is calculated by the proportion of negative values in the distribution of the difference.

The results of this sequence of procedures for the three separate studies are summarized in Tables 1 to 6. Attention in the analyses of efficiency effects is focused, however, on the Escanaba and the Sandhill studies, as they adequately demonstrate the various effects of joint estimation and cross equation restrictions. Descriptive statistics and definitions of the variables used in the maximum likelihood estimates of the univariate and bivariate probit models for these studies are provided in Table 1. Following Boyle, Welsh, and Bishop, analyses of the *Grand Canyon White Water Boaters* survey responses involved simple models with the only covariates being the cost of the actual trip taken and the bid value for the hypothetical flow scenario.[6]

Table 2 summarizes the independent and joint estimation results for the "baseline permit" and the "15 percent fewer" valuation questions asked in the Escanaba fishing study. The first column presents independently estimated valuation functions. The second and third columns present the joint unrestricted and restricted models respectively. Although varying in significance, the signs of the estimated coefficients are consistent across equations: the probability of a "yes" response increases with perceived importance of the resource, distance traveled to Lake Escanaba, and the educational level of the respondent, but falls with increasing bid values. Importantly, there is an extremely high correlation (ρ=0.92) in estimated response functions across the two dichotomous choice CV questions. Likelihood ratio tests demonstrate that this correlation coefficient is highly significant. Casual comparison of estimated parameters suggests that the WTP response functions are quite similar across scenarios.

A likelihood ratio test of H_0^2: $\beta_{\text{Baseline Permit}} = \beta_{\text{15 Percent Fewer}}$ is rejected at the 5 percent significance level (LR = 11.364, $\chi^2_{5,\,0.05}$ = 11.05), implying that the response functions are different in spite of the fact that they are significantly and highly correlated. The final restricted model was arrived at by testing various individual and joint cross-equation restrictions for coefficients. For this data set, the equality restrictions hold for the estimated coefficients for the Import, Miles 1, and Bid variables. The hypothesis of cross-equation equality for the Education coefficient was rejected. In the baseline scenario equation this coefficient was positive and significant, but was not significant for the 15 percent fewer model. The cause of this difference across equations is not identified.

Inspection of the asymptotic standard errors in each of the models indicates that there is little efficiency gain from estimating the joint model without restrictions. Indeed, the asymptotic standard errors on some of the coefficients actually increase with joint estimation, and the significance of the coefficients on Education (baseline permit), and the Miles 1 (15 percent fewer) crossed standard significance level thresholds. In contrast, a slight efficiency gain is noted for all variables as a result of imposing the cross equation restrictions. However, for the most part, the efficiency gains from imposing these restrictions are negligible.

Independent and joint estimation results for the "either sex" and the "antlerless" Sandhill hunting permits are provided in Table 3. The probability of a "yes" response increases with the perceived quality of the hunting experience, but declines with higher dichotomous choice posted prices. The coefficient on the year variable was only significant for the antlerless model, indicating that the 1992 respondents had higher values for the antlerless permits. This result is consistent with the observation that the 1992 Sandhill hunt was limited to antlerless deer, and that the respondents generally reported a positive experience in spite of the fact that an antlerless hunt is popularly regarded to be inferior to an either sex hunt. Although the estimated correlation coefficient of 0.39 is much lower than the Escanaba study, it is still highly significant, indicating that substitution effects, if they exist, do not offset factors favoring a positive correlation. This lower correlation is reflected by the observation that there is an obvious difference in parameter estimates across scenarios. Notably, the effect of prices on WTP is more distinct for antlerless permits, suggesting both a lower value and variance in values for WTP in the antlerless scenario.

All possible combinations of individual and joint coefficient restrictions across the sandhill equations were rejected using likelihood ratio tests with the unrestricted model as a reference. This demonstrates that entire valuation functions can be significantly different even though a significant correlation across equations is observed. Like the Escanaba study, the asymptotic standard errors of the joint-unrestricted model are quite similar to those of the independent model -- indicating little efficiency gains from joint estimation.

The individual, the joint-unrestricted, and the joint-restricted models were estimated for each of the six possible *Grand Canyon White Water Boating Survey* valuation comparisons (5 vs 13 kcfs, 5 vs 22 kcfs, 5 vs 40 kcfs, 13 vs 22 kcfs, 13 vs 40 kcfs, and 22 vs 40 kcfs)[7]. In two comparisons (5 vs 13 kcfs and 5 vs 22 kcfs) correlation across equations was rejected, and the two valuation response functions are statistically independent. Comparison of flow descriptions suggests that this result is not surprising, as the description of the 5 kcfs low flow scenario differs considerably from those at the more desirable moderate levels. As noted previously some of the negative attributes of the 5 and the 40 kcfs scenarios were similar, which is consistent with the result that the correlation coefficient between WTP functions for these two scenarios was significant at the 10 percent level. All other pairwise correlations were significant at the 1 percent level.

Rejection of cross-equation equality restrictions also varied across the four pairwise comparisons that have significant correlation coefficients. The joint hypotheses that all coefficients were equal could not be rejected at the 1 percent level for the 13 vs 22 kcfs comparison, indicating that these flows have statistically similar valuation functions when correlation is accounted for. Hypotheses of equality across equations of the bid coefficient, the bid and constant coefficient, and all but the bid and cost coefficients could not be rejected for the 5 vs 40, the 13 vs 40, and the 22 vs 40 kcfs pairwise comparisons, respectively. Again, in spite of significant correlation and the failure to reject selected cross-equation coefficients, efficiency effects on the individual coefficients were found to be minor when moving from the independent to the joint models.

Taken together, the results presented so far demonstrate that there can be significant correlation between responses to contingent valuation questions elicited in the same survey. There are some indicators that this correlation is quite high when the attributes of the commodity being valued are quite similar across questions and declines with dissimilarities. In spite of the fact that the attributes vary widely across scenarios, the sign of the correlation coefficient is either positive and significant, or not significantly different from zero. Given that some scenarios encompass very different attributes, the lack of any negative correlation coefficients suggests that systematic effects outweigh substitution effects[8].

Table 5 provides summary statistics for the approximate WTP distributions estimated for the individual and joint models for cases where the correlation coefficient was significant at the 10 percent level or better, and joint equality restrictions across equations for all elements of the coefficient vectors could not be rejected. The first column provides the correlation coefficient. The next three columns identify the distributions for the first scenario being compared and provide the bootstrap results from the joint and independent estimations of mean WTP using the "best" joint-restricted model in which cross equation equality restrictions cannot be rejected. The fifth through seventh columns present the same information for the second scenario. The final column provides a ratio of the variance of the distribution of the differences from the joint model ($\sigma^2_{x\text{-}y,\ JOINT}$) to the variance of the difference from the independent model ($\sigma^2_{x\text{-}y,\ INDEP}$). This relationship is of particular interest because positive correlation in mean WTP values is expected to reduce the variance of the difference, as suggested in the introduction.

A comparison across columns in Table 5 shows that even when the correlation coefficient is relatively high, there is only a very small effect on the estimated distributions of mean WTP associated with estimating joint models. Confidence ranges change only slightly, if at all, between the independent and joint models, suggesting that efficiency effects on the variance of the difference are minor[9]. This result is consistent with the almost negligible effects of joint modeling on the values of the estimated coefficients observed in Tables 2 and 3. However, the ratio of the joint to independent variance of the difference does decline with increases in the level of correlation. Combined, these results suggest that as correlations rise, the distribution of the difference estimated under the independence assumption will generally tighten up even though there may be no observed effect on individual distributions.

Table 6 indicates that accounting for the correlation in the estimation process does impact on difference of means tests for comparisons in which the significance of the difference fell in statistically interesting ranges (i.e., around 0.10, 0.05, 0.01). Values associated with other comparisons in Table 4 diverged substantially from these critical values (e.g., 0.45 or 0.000000001) and, thus, are not that interesting in terms of this analysis. The values provided in Table 6 demonstrate that accounting for the correlation does have an effect on the decision of whether to accept or reject the hypothesis that the mean WTP is significantly different across scenarios. In each case presented it causes the estimated significance values to cross the 5 percent level, and, in the 13 kcfs - 40 kcfs comparison, the critical significance level changes from 0.1 to 0.01. As such, joint estimation appears to have potentially important consequences from a policy perspective.

4. IMPLICATIONS AND CONCLUSIONS

The high correlation coefficients observed and the consequent effects on the difference in means distributions indicate that across scenario correlation may be an important factor in some policy comparisons, and is particularly high in closely related and embedded scenarios. As such, this paper provides empirical evidence from actual CV studies that standard assumptions of independence in comparing distributions lead to biased estimates of the difference, and may lead to erroneous conclusions about the significance of the difference in mean WTP values elicited in the same questionnaire.

From an applied perspective this empirical result must be weighed against the additional programming costs of implementing the joint estimator. If the cost of adopting the more complex bivariate probit approach is perceived to be high for the individual researcher, it is critical to recognize that there are instances in which the additional programming costs might not be warranted from a difference of means perspective. For example, a rule of thumb might be to not use joint estimation if independent mean WTP distributions overlap considerably, say at the 20 percent level or higher. Under these conditions it is unlikely that joint estimation will change the decision to not reject the null hypothesis of equality. At the other extreme, distributions that do not overlap at all when estimated independently, would indicate that -- unless there was a strong reason to believe that responses are highly negatively correlated -- the bivariate probit estimation approach would not change the hypothesis test results. At the same time, it should be acknowledged that the additional costs of joint estimation should not be prohibitive. Standard statistical packages such as LIMDEP have readily accessible bivariate routines.

The relative cost-benefit ratio of imposing cross equation restrictions is much larger. Cross equation equality restrictions are not an option in most standard statistical packages, and thus the researcher must possess more sophisticated programming skills in order to impose these restrictions[10]. As demonstrated in the text, the effect of these restrictions on the standard errors of individual parameters and on mean WTP distributions is slight. Any efficiency gains on the difference of means test appears to be dominated by correlation effects. In projects where the research objective is not a difference of means test between scenarios (such as instances where disparate valuation questions are elicited) the results from this paper further suggest that there is little benefit from conducting a joint estimation procedure. This latter result is in line with findings presented in Alberini and Kanninen for contingent valuation, but conflict with prior expectations based on previous econometric results in other applications. Given these conflicting results, it might be premature to conclude that efficiency gains from joint estimation and cross equation restrictions are negligible.

In all, more empirical research is warranted before conclusions could be drawn about the importance of efficiency effects. Similarly, the source of correlation is not isolated in this study, and future research should also be directed towards identifying whether correlation is attributable to perceived similarities in attributes across scenarios, or to a number of psychological response factors that have recently been suggested in the literature.

Table 1. Variable Definitions and Descriptive Statistics for Escanaba and Sandhill Studies

Escanaba	Description	Mean (s.d.)
Import	Categorical response "If I could not go fishing at Escanaba Lake during the 'early season', I would": 1) easily find something else to do; 2) miss it, but not as much as other things that I enjoy; 3) miss it more than the other interests I now have; 4) miss it more than all the other interests I now have.	2.01 (0.89)
Miles	Open ended variable: distance between Escanaba Lake and home (one way).	120.41 (115.05)
Education	Categorical response: 1) less than high school; 2) high school graduate; 3) some college or technical school; 4) technical or trade school graduate; 5) college graduate; 6) advanced degree.	3.20 (1.47)
Bid 1	Dichotomous choice bid value for "baseline permit".	13.60 (12.43)
Bid 2	Dichotomous choice bid value for "15 percent fewer".	11.34 (9.62)
Sandhill		
Quality	Categorical response for quality of the hunt: 1) very low quality; 2) fairly low quality; 3) average quality; 4) fairly high quality; 5) very high quality.	3.40 (1.23)
Year	Binary variable: 1991=1, 1992=2.	1.51 (0.50)
Bid 1	Dichotomous choice bid value for "either sex" permit.	27.91 (21.81)
Bid 2	Dichotomous choice bid value for "antlerless" permit.	29.12 (24.58)

Table 2. Escanaba Fishing Study[a,b,c]

	Independent	Joint, Unrestricted	Joint, Restricted
Full permit			
Constant	-0.5860 (0.2492)**	-0.5090 (0.2420)**	-0.5100 (0.2335)***
Import	0.2172 (0.0777)***	0.2488 (0.0767)***	0.2609 (0.0656)***
Miles 1	0.0014 (0.0006)**	0.0016 (0.0006)**	0.0013 (0.0005)***
Educ	0.1356 (0.0489)***	0.1239 (0.0508)**	0.1286 (0.0483)***
Bid 1	-0.1229 (0.0131)***	-0.1350 (0.0100)***	-0.1371 (0.0079)***
15 percent fewer			
Constant	-0.3371 (0.2522)	-0.3166 (0.2573)	-0.3082 (0.2344)
Import	0.2685 (0.0760)***	0.2738 (0.0780)***	See Import, full permit
Miles 1	0.0012 (0.0006)**	0.0011 (0.0006)*	See Miles 1, full permit
Educ	0.0208 (0.0466)	0.0241 (0.0474)	0.0186 (0.0472)
Bid 2	-0.1346 (0.0141)***	-0.1390 (0.0126)***	See Bid 1, full permit
ρ		0.9173 (0.0426)***	0.9160 (0.0354)***
Likelihood ratio χ^2_1	180.79		
Likelihood ratio χ^2_2	142.32		
-Log Likelihood[b,c]	-218.63 - 228.02	-391.66	-391.94
n	540	540	540

[a] Numbers in ( ) are asymptotic standard errors.
[b] *, **, *** indicate significance levels of 0.10, 0.05, and 0.01 respectively.
[c] $-2\,(LL_i - LL_{j,u}) = 110.04,\ \chi^2_{1,0.10} = 2.71$

Table 3. Sandhill Deer Hunting, 1991-1992[a,b]

	Independent	Joint, Unrestricted
Either sex		
Constant	0.9977 (0.4245)**	0.9919 (0.4349)***
Quality	0.1224 (0.0871)	0.1278 (0.0876)
Year	0.1361 (0.2052)	0.1266 (0.2074)
Bid 1	-0.0369 (0.0056)***	-0.0371 (0.0057)***
Antlerless		
Constant	-1.378 (0.3919)	-0.1008 (0.3941)
Quality	0.2321 (0.0922)***	0.2288 (0.0933)***
Year	1.1225 (0.2197)***	1.1245 (0.2323)***
Bid 2	-0.0704 (0.0129)***	-0.0732 (0.0118)***
ρ		0.3923 (0.1382)***
Likelihood ratio χ^2_1	60.52	
Likelihood ratio χ^2_2	82.73	
- Log Likelihood[c]	-105.98 - 93.32	-197.47
n	197	197

[a] Numbers in ( ) are asymptotic standard errors.
[b] *, **, *** indicate significance levels of 0.10, 0.05, and 0.01 respectively.
[c] $-2\ (LL_i - LL_{j,u}) = 7.35,\ \chi^2_{1,0.10} = 2.71.$

Table 4. Correlation Coefficients Across Bivariate Probit Models for Different Flow Levels: Private Boaters[a,b]

	5 kcfs	13 kcfs	22 kcfs
13 kcfs	0.23 (0.23)		
22 kcfs	0.27 (0.18)	0.87 (0.08)***	
40 kcfs	0.32 (0.17)*	0.84 (0.16)***	0.63 (0.17)***

a Numbers in ( ) indicate asymptotic standard errors.
b *,**,*** indicate significance levels of 0.10, 0.05 and 0.01 respectively.

Table 5. Individual and Jointly Estimated Mean Value Distributions[a]

	Distribution 1			Distribution 2			$\sigma^2_{x-y,JOINT}$
ρ	Name	Mean (Indep)	Mean (Joint)	Name	Mean (Indep)	Mean (Joint)	$\sigma^2_{x-y,INDEP}$
0.32	40 kcfs	432 [369, 505]	431 [372, 493]	5 kcfs	243 [198, 296]	237 [185, 298]	0.84
0.39	Sandhill_{ES}	40.17 [36.88, 45.51]	40.85 [36.64, 45.15]	Sandhill_{A}	18.08 [15.75, 21.86]	17.72 [15.43, 21.08]	0.80
0.63	22 kcfs	522 [454, 599]	527 [454, 596]	40 kcfs	432 [369, 505]	420 [352, 491]	0.74
0.84	13 kcfs	518 [458, 588]	528 [466, 592]	40 kcfs	432 [369,505]	417 [347, 480]	0.36
0.92	$\text{Escanaba}_{Baseline}$	5.40	5.51	$\text{Escanaba}_{15\%\ fewer}$	4.74	4.74	0.52

a Numbers in [ ] reflect the 0.90 confidence interval.

Table 6. Significance Levels of Difference of Mean WTP Estimates: Selected Observations

ρ	Comparison	$\hat{\alpha}_{indep}$	$\hat{\alpha}_{joint}$
0.63	22 kcfs - 40 kcfs	0.066	0.026
0.84	13 kcfs - 40 kcfs	0.054	0.003
0.92	Escanaba	0.086	0.016

REFERENCES

Alberini, A., 1995. "Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models," *Journal of Environmental Economics and Management,* 29(2):169-180.

Alberini, A. and B. J. Kanninen, 1994. "Efficiency Gains from Joint Estimation: When Does a Second Equation Improve Estimation of the First?," paper presented at the AERE/AAEA annual meetings, San Diego.

Boyle, K. J., M. P. Welsh, and R. C. Bishop, 1991. "An Empirical Investigation of the Role of Question Order, Symbolic Effects, and Respondent Experience in Contingent-Valuation Studies", Unpublished manuscript, Department of Agricultural and Resource Economics, University of Maine, Dec.

Boyle, K. J., M. P. Welsh, and R. C. Bishop, 1993. "The Role of Question Order and Respondent Experience in Contingent-Valuation Studies", *Journal of Environmental Economics and Management*, 25(1):S80-S99.

Cameron, T. A., 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censured Logistic Regression," *Journal of Environmental Economics and Management*, 15:355-379.

Cameron T. A., and M. D. James, 1987. "Efficient Estimation Methods for "Closed-Ended" Contingent Valuation Surveys," *Review of Economics and Statistics*, 69(2):269-276.

Cameron, T. A. and J. Quiggin, 1994. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-Up' Questionnaire," *Journal of Environmental Economics and Management,* 27(3):218-234.

Champ, P. A. and R. C. Bishop, 1996. "Evidence on the Accuracy of Expenditures Reported in Recreational Surveys," *Journal of Agricultural and Resource Economics*, forthcoming July.

Goldberger, A. S., 1991. *A Course in Econometrics*. Harvard University Press, Cambridge.

Greene, W. H., 1993. *Econometric Analysis*, 2nd Ed., Macmillan Publishing Company, New York.

Hanemann, W. M., 1984. "Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses", *American Journal of Agricultural Economics*, 66(3):332-341.

Hanemann, W. M., 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses: Reply", *American Journal of Agricultural Economics,* 71(4):1057-1061.

Hsiao, C., 1986. *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Kahneman, D. and J. L. Knetsch, 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction", *Journal of Environmental Economics and Management*, 22(1):57-70.

Kanninen, B. J, 1995. "Bias in Discrete Response Contingent Valuation", *Journal of Environmental Economics and Management*, 28(1):114-125.

Krinsky, I. and A.L. Robb, 1986. "On Approximating the Statistical Properties of Elasticities", *Review of Economics and Statistics*, 68(Nov):715-719.

Mood, A., F. Graybill, and D. Boes, 1974. *Introduction to the Theory of Statistics*, McGraw Hill, San Francisco.

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.

Poe, G. L., E. K. Severance-Lossin, and Michael 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. Blomquist, 1995. Contingent Valuation When Respondents are Ambivalent, *Journal of Environmental Economics and Management*, 9(2):181-196.

Appendix

Text of Escanaba Lake Survey "Full Permit"

28. Suppose that a special permit will be required this year to fish Escanaba Lake during the "early season". Assume that you can order a permit to fish the "early season" at Escanaba Lake by mail. A permit is valid for the two weeks before the open of the regular fishing season (April 17 through May 1, 1992).

    - All individuals 16 years old and older wishing to fish Escanaba Lake between April 17 and May 1, 1992 will have to show a permit at the research contact station.
    - Individuals with a permit may fish Escanaba Lake as often as they would like between April 17 and May 1, 1992.
    - All the regulations currently in force for fishing Escanaba Lake will stay the same as they are now.
    - If the ice goes out early, fishing will be free until April 17.
    - The revenue from permit sales will go to the Northern Highland Fishery Research Area.

The amount we ask about below may seem very high or low to you but it's very important that you answer the question seriously. The amount written below was randomly assigned to you.

Would you pay $_____ for a permit to fish Escanaba Lake between April 17 and May 1, 1992? (CIRCLE ONE NUMBER)

1. No
2. Yes

29. Assume that permits would be sold as described in question 28. In addition, assume that there will be about 15 percent fewer walleye in Escanaba Lake than usual at the beginning of the "early season". The size of the fish would be the same as now. It is hard to say how this would affect the catch of any one angler. Some anglers may catch just as many fish as usual while others may not do as well. In thinking about how this might affect your success, assume that there will be somewhat fewer walleyes around. As in question 28, the amount written below was randomly assigned to you.

    Under these new conditions, would you pay $_____ for a permit to fish Escanaba Lake between April 17 and May 1, 1992? (CIRCLE ONE NUMBER)

    1. No
    2. Yes

Text of Sandhill Public Deer Hunt Survey (1991)

16. Suppose that next year you apply for a Sandhill General Public Deer Hunt permit but are not chosen to receive a permit. Imagine that as part of a research project you have the chance to purchase an either sex permit. If you were able to buy a 1992 Sandhill either sex permit, would you be willing to pay $_____? (CIRCLE ONE NUMBER)
    1. No
    2. Yes

17. If you were able to buy a 1992 Sandhill antlerless permit, would you be willing to pay $____? (CIRCLE ONE NUMBER)
    1. No
    2. Yes

Text of Grand Canyon White Water Boater Survey
22 kcfs

At moderately high water levels (around 22,000 cfs), the pace of the river is faster than at lower flows, leaving more time for side canyons and stops at attractions. Boating groups do not have a problem staying on schedule. Rapids have larger waves and provide a bigger "roller coaster" ride than at moderate water. Only a few passengers choose to walk around some of the bigger rapids for their safety. Some potential campsites are under water in some areas of the canyon, but generally campsites are plentiful although a bit smaller in size.

We would now like you to image that you are presently deciding whether or not to go on a Grand Canyon white water trip. Imagine that the trip would be the same as your last trip (e.g. the people, food, etc.) with two exceptions:

The water level would be constant at 22,000 cfs (see description for Case 4 above)

AND

Your individual costs for the trip increased by $_____ (over the total cost you calculated on page 8, question A26)

D2. Would you go on this trip? (CIRCLE ONE NUMBER)
1. YES, I WOULD PAY THIS AMOUNT TO TAKE THE TRIP
2 NO, I WOULD NOT PAY THIS AMOUNT TO TAKE THE TRIP

Notes:

1. Papers by Park, Loomis, and Creel, and Ready, Whitehead, and Blomquist provide examples of this format.

2. This point was raised by an anonymous reviewer. The reader is referred to a related paper by Alberini and Kanninen which explores the efficiency gains associated with joint estimation and cross equation restrictions of various combinations of continuous and discrete contingent valuation response formats.

3. The motivation and conceptual framework parallels that used to support bivariate probit models in double bounded dichotomous choice contingent valuation detailed in Cameron and Quiggin and Alberini.

4. Other currently used techniques for comparing empirical distributions such as the non-overlapping confidence interval criterion [e.g. Park, Loomis, and Creel] or normality assumptions [e.g. Desvousges *et al.*] are biased or are otherwise not appropriate for general applications [Poe, Severance-Lossin, and Welsh]. Analytical solutions [e.g. Cameron, 1991] for median WTP values in the DC framework offer another possible approach, but are not generalizable to situations such for which analytical solutions do not exist. Moreover, empirical bootstrapping approaches have been shown to approximate analytical solutions in general [Efron and Tibshirani] and for dichotomous choice contingent valuation in particular [Ballestreri *et al.*].

5. Unfortunately, the actual WTP data was never collected.

6. An appendix of the WTP distribution estimates of the *Grand Canyon White Water Boaters Survey* is available from the authors.

7. As noted by an anonymous reviewer and W-133 participants, a more complete model would estimate all four scenarios simultaneously. Instead, bivariate probit models were estimated for each of the pairwise comparisons in this analysis. In addition, some caution should be taken in interpreting the magnitude of the correlation found in this analysis because the original study involve possible effects associated with different orderings of the contingent valuation questions (see Boyle, Welsh, and Bishop, 1991, 1993). High correlations between scenarios may be attributed to proximity in the survey rather than to similarity in conditions. Similarly, low correlation values might be attributed to ordering effects. In spite of these limitations, the paired comparisons and the data are retained in this manuscript for illustrative purposes.

8. That substitutability and negative correlation across question responses can occur in a random utility bivariate probit framework is demonstrated in a study by Horowitz (1994) of consumer preferences for government programs that reduce anthropogenic risks. In that study, however, substitutability was forced by explicitly altering the relative risks of competing programs and relegating the number of lives saved assumed by respondents (which are not observable) into the error term.

9. Across the ten possible comparisons, the independent to joint ratio of the variance of mean WTP distributions averaged 0.9531 (0.1550) but was not significantly different from 1 ($t=0.467$).

10. Programming in this paper was conducted in Gauss 3.11, using the maxlik application.

REFERENDUM VOTING STRATEGIES AND IMPLICATIONS FOR FOLLOW-UP OPEN-ENDED RESPONSES

Michael C. Farmer, Asst. Prof.
School of Public Policy
Georgia Institute of Technology

Alan Randall, Prof.
Dept. of Agricultural Economics
Ohio State University

ABSTRACT

A new set of diagnostics are introduced to test the consistency between dichotomous choice and open-ended follow up responses in willingness to pay surveys. If conformity between the discrete choice data and the open-ended follow-up data adhere to a strict set of regularity conditions, then valid WTP inferences can be extracted from the open-ended responses. That conformity predicts a specific relationship exists between open-ended responses and starting points. Since voters may react conservatively to the costs of services in actual referenda, the additional information provided by the follow-up response assesses the presence and magnitude of any conservative responses.

Correspondence to: Dr. Alan Randall, Prof; Dept. Agr Econ.
Ohio State University. Columbus, OH 43202

INTRODUCTION

Dichotomous Choice questions have emerged as the premier demand elicitation technique in willingness to pay (WTP) surveys. Respondents are asked to make a dichotomous choice to either accept or to reject a proposed service at a given price. The process has been linked to voting in public referenda. The referendum survey format boasts several advantages over a direct open-ended WTP elicitation format. Principally the referendum design presents the respondent with a take-it-or-leave-it (TILI) proposition - a structure widely considered immune from strategic manipulation. The one-shot finality of the choice motivates an honest revelation of preferences [17]. Flanked by continued advances in latent dependent variable estimation, honest revelation of a weak inequality promises both an unbiased and efficient WTP estimate [6, 14, 7].

Nonetheless single dichotomous choice (DC) practice has been frustrated by efficiency concerns. To add information to the elicitation of a weak preference inequality, innovations in DC survey formats began to collect follow up DC responses. These double-bounded formats retain the basic YES/NO referendum structure and introduce substantially more information into the estimation process [15]. Results from this survey design, however, suggest that the behavior of the first DC response differs from that of the second. Studies show that follow up DC responses are from different distributions [8], due perhaps to response anchoring [16], to changing preferences [1] or to response strategies entirely separate from TILI.

This problem is not new to CVM research. The concern echoes problems of suggested price (SP) effects long identified with open-ended responses following referenda questions [25, 5]. SP effects mean that open-ended responses which follow DC questions are strongly correlated with the value of the initial DC price. In some respects the research focus has turned full circle. Boyle et. al. [5] directly identify pervasive starting point effects in continuous data, underscored by a skeptical view to then current anchoring corrections, as part of their rationale for adopting DC data for WTP estimation. Current suggestions, however, now adapt Thayer's [25] original form of anchoring correction to the statistical structure of the DC follow-up [16].

Efficiency frequently mandates some follow up response in WTP surveys. Yet follow up responses exhibit patterns of bias in both their discrete and continuous varieties. This demands a behavioral method to reconcile the first and second responses to one another. Given the experience of referendum survey practice, it also appears that persuasive SP corrections out of basic theory for the "biases that might be introduced" [2] promise considerable efficiency gains in WTP estimation.

To reconsider starting point effects, we extend the discussion of suggested price effects beyond anchoring [24]. This requires a review of the incentive compatibility of DC responses introduced by Hoehn and Randall [17]. Therefore we first return to the Public Choice literature from which the Hoehn and Randall results are partly drawn and explicitly detail the strategic motives of the referendum voter consistent with the findings of that literature.

A fully satisfactory indicator of WTP from a DC survey is consistent with a successful budget maximizing public agency. That agency must convince voters that proposed referenda are true one shot take-it-or-leave-it (TILI) propositions. Yet the success of TILI strategies by budget maximizing agencies is typically incomplete. This is because a wholly satisfactory response commits the voter to a complete exhaustion of economic surplus whenever the suggested referendum price equals their WTP. Some degree of bargaining by the ballot box compels agencies to suggest referenda prices strictly below maximum WTP for the median voter. This implies that referenda voting is a conservative indicator of WTP, at least from this effect. Referendum voters respond so as to realize positive expected surplus gains through their referendum responses, a behavior formalized herein.

The theoretical explication of surplus maximizing voting strategies is then extended to the open-ended follow up question. If continuous responses are consistent with voting behavior, it is shown that a fairly rigid set of predictions about the open-ended data are inferred. More importantly under some restrictions, valid WTP inferences can result directly from the continuous follow up response. What emerges is an empirical procedure to diagnose the incentive behavior of a CVM survey.

To illustrate this WTP inference process and the internal consistency checks offered, an empirical illustration is presented. Results fail to reject the voting strategy outlined by this study. Yet the choice of suggested prices (SPs) reflects a preference for DC estimation that, consistent with this theory, pre-empts a convincing single WTP point estimate.

PUBLIC CHOICE AND TAKE IT OR LEAVE IT REFERENDA

Reconciling incentives in dichotomous choice WTP surveys to voting behavior in public referenda is accomplished largely by assertion. Richard Carson [9] did conduct an important test of a CVM instrument against a real California referendum with some success. Yet generalizations are restricted by the range of goods over which WTP surveys are applied and by the range of payment vehicles adopted. Typically we won't have public elections to check our results; so we must either be satisfied to assert a take-it-or-leave-it reaction to a WTP survey or to look elsewhere within the survey for evidence consistent with that behavior.

Prior to offering any specific design features to uncover respondent motives, we need to be explicit about the voting response incentives expected in public referenda.

Both empirical and experimental evidence on referendum voting does locate take-it-or-leave-it reactions by voters; but it is found within a larger set of payoff maximizing strategies.

Romer and Rosenthal [20] first hypothesized that agencies could secure larger budgets by exploiting informational advantages regarding the costs of service provision and by controlling the timing of elections. The authors hypothesized that referenda would establish budget levels well above median voter hypothesis levels to levels just passably acceptable to a voting majority. The bureau's advantage in this game has been formalized as asymmetric information [12]. Since agencies uniquely understand the policy process and the range of options for providing the good in question, they can manipulate the budget setting process to their advantage. The full agenda setting budget level is realized whenever voters approve public service provision at costs equal to the maximum WTP of voters at the 50th percentile demand level. This of course exhausts all economic surplus gains from the proposal for the median voter. Such a budget exceeds the level supported by the median voter hypothesis which, conversely, assumes budget levels maximize economic surplus of the median voter contingent on the array of public goods available.

Romer and Rosenthal conducted a series of empirical investigations into School Board referenda in Oregon. Their results find that local school boards determine the timing and size of bond issue referenda to secure budgets well beyond that suggested by the median voter hypothesis [21, 22].

Complete agenda setting success directly implies that referenda results represent a wholly satisfactory indicator of voters' preferences for provision of the good. Voters approach the referendum tax price as a credible take-it-or-leave-it offer and accede to the tax price until it surpasses their maximum WTP. There is a theoretical basis for presuming such budget maximizing strategies will be largely successful. Implementation tools to limit a bureau's take-it-or-leave-it agenda setting power [3] are restricted for single ballot propositions.[1] This is not only because voters are diffused. Rather the referendum structure prevents voters from easily bargaining with the agency.

Nonetheless applied works found that the voting public is not wholly subordinate to the budget extremes of the pure agenda setting agency. Bargaining does apparently affect final budget solutions somewhere between the median voter hypothesis level and the extreme agenda-setting agency level.

The first convincing test was established experimentally [10]. Voters it seems respond to the prospect for repetition in the game; and they anticipate repetition even before receiving a specific follow up cue. A later work [22] found that anticipated prospects for follow up introduce a degree of

1. An example of a suggested principle-agent solution has the principle adjust agency output at the unit price suggested. For the referendum suggested price this power seems absent as a single proposition names a fixed price *and* a fixed quantity supplied.

bargaining discipline. This moderates costs of government service away from complete agenda setting levels. The presence of this more general voting strategy compromises assumptions that ground the strong take-it-or-leave-it (TILI) game forms required to achieve a wholly satisfactory indicator of WTP from referendum responses alone.

Complicating the issue further, evidence suggests that the degree of agenda-setting success varies over time and over type of service. Issues decided by referenda where delay costs disadvantage both the agency and the public (a hospital construction bond issue for instance) result in budgets more toward the median-voter hypothesis level and away from the TILI hypothesis level [13]. Voters understand that defeated issues can re-appear; and this threat is not lost on the bureaus. Agencies that fail to set suggested referendum prices strictly below the TILI levels greatly risk rejection. Timing is also important. The time horizon over which budgets are negotiated and voter expectations are formed can be quite long. Shapiro & Sonstelie [23], for example, interpret California's notorious Proposition 13 as a long run, global play in this bargaining game that pulled budgets back from more surplus exhausting extremes of the agenda-setter. The contingent nature of voting militates against easy generalizations about the incentives affecting a specific vote. If contingent valuation surveys reflect this strategy, those surveys can be expected to exhibit at least this variability.

We conclude that voters balance the expected loss of delay (including the risk of permanent defeat) against expected gains of policy implementation when casting their votes. Voters attribute these expectations issue by issue and period by period depending on the importance of the question, its expense and their experience with a given agency's pricing practices. Although no pre-determined prediction for a given vote avails itself, some general principles are robust. Cost differences indicate that public agencies provide the same level of service at consistently higher prices than do private suppliers, meaning budgets lie somewhere above the unqualified median voter hypothesis [4]. In general, evidence supports a claim that budgets lie somewhere between the median voter hypothesis level and the surplus exhausting level of the TILI budget-maximizing agenda setter [19].

If we want to suggest a model for extending referendum voting with the prospect for voter bargaining to a CVM survey, we need to detail a formal structure for the voting model.

STRATEGIC VOTING

To review the implications of this voting strategy on referendum responses, we characterize proposals suggested by referenda. The baseline condition for the respondent is represented by the expenditure function:

$$e(P,Q^o,U^o) = e^o.$$

The referendum offers a public service improvement, so:

$$e(P,Q^1,U^o) = e^1.$$

The Willingness to Pay for the change in baseline service is:

$$e(P,Q^o,U^o) - e(P,Q^1,U^o) = WTP.$$ [2]

The referendum advertises this service level improvement at a suggested tax price (SP). Economic surplus from realizing this change in the state of the world is defined by:

$$WTP - SP = \Delta CS.$$

If the voter is a surplus maximizer, the referendum question becomes:

IS $$EU\ (YES) \geq EU\ (NO)\ ?$$

So we can now consider the payoff to referendum voting under the behavior discussed.

YES VOTES: Results from referenda are not wholly decisive events. Some uncertainty exists in the effectiveness of YES votes to resolve a policy question.

A YES vote in a referendum generates the expected payoff:

2. For those unaccustomed to using the expenditure function, it is useful to keep in mind that a positive WTP, meaning the service improvement is valued, implies that $e^1 - e^o < 0$.

$$\text{Expenditure(YES)} = [1 - \pi(SP)] * [e^1 + SP] + \pi(SP) * e^o.$$

where $\pi'_{SP} < 0$. $\pi''_{SP} > 0$. $\partial^2 \pi / \partial Bid \partial SP > 0$.

$(1- \pi)$ is the respondent attributed probability that the referendum process is policy decisive: YES means the proposal actualizes. Referenda that pass yet fail to change policy generate payoffs equal to the baseline expenditure (e^0) with probability π. Initially, π contains only SP as an argument. Larger SPs signal to the respondent that acceptances are more likely decisive to a budget maximizing agency; so, π declines as SP increases. If the proposal carries and becomes policy, the referendum tax represents the personal costs of acquiring Q^1. So when the referendum is decisive, it reflects the *ex post* change in consumer surplus: $e(P,Q^o,U^o) - e(P,Q^1,U^o) - SP$ (i.e. WTP - SP). These gains decline as SP grows. Yet as decisiveness itself occurs with probability $[1- \pi(SP)]$, this probability rises as SP grows.

NO VOTES: A No vote generates a similar expected payoff structure. Yet rejection opens the door for a defeated proposal to be re-considered at a later date. Though any strict utility improving re-consideration inside the $\{Q,SP\}$ set is permitted, for simplicity, improvements are cast as simple price changes for a given Q^1-Q^o program. Expected improved reconsiderations are, therefore, representable by a probability distribution over the dimension of costs (P).

A NO vote in a referendum generates the expected payoff:

$$\text{Expenditure(NO)} = \int_0^{SP} pdf(P{:}SP) * [e^1 + P]\, DP + e^o * \{1 - \int^{SP} pdf(P{:}SP)\, DP\}\ \}.$$

The subjective expectations of lower prices are distributed along a respondent attributed probability density function, pdf (P:SP), from zero to the referendum price. The pdf reflects the probability that a given counter-price eventually becomes policy following referendum rejection. So the cumulative distribution, $\int^{SP} \{pdf(P{:}SP)\}\, DP$, is strictly less than unity and the CDF is discontinuous in price space. Discontinuity occurs because there remains a fixed probability that policy rejection permanently defeats the issue. A NO vote invites expenditures, $e^1 + P$, under the cumulative probability $\int^{SP}\{pdf(P:SP)\}\ DP$. e^o is realized by permanent rejection with probability $1 - \int^{SP}\{pdf(P:SP)\}\ DP$.

INDIFFERENCE: The voter resolves the acceptance or rejection dichotomy by locating a critical indifference SP, called SP_c, where expected payoffs of a YES and NO vote are equal.

Consider first the take-it-or-leave-it (TILI) indifference price in the literature. In this case the chance that the issue will be revisited is presumed zero. The voter assumes:

TILI: $$\int_0^{SP}\{pdf(P:SP)\ DP = 0.$$

More generally, the pdf is simply very dense on SP. This means the TILI pdf realizes its entire value at P = SP. Therefore, EU(YES) = EU(NO) is defined by:

$$[1 - \pi(SP_c)] * [e^1 + SP_c] + \pi(SP_c)*e^o = e^o.$$

Or, $$SP_c = WTP$$

and $$\Delta CS = 0.$$

Voters are then indifferent only when tax prices exhaust all potential utility improvement of the service proposal. If $\int_0^{SP}\{pdf(P:SP)\ DP = 0$, the referendum is a strategically satisfactory indicator of WTP. Specifically, it is a dominant strategy to:

Reject if SP > WTP,

Accept if SP < WTP,

be Indifferent if SP_c = WTP.

The Hoehn and Randall [17] conclusion then can be generated as a special case of the referendum voting behavior outlined. The chance of re-visitation, however, may be positive. So a more general case implies:

$$\int_0^{SP}\{pdf(P:SP)\ DP \geq 0.$$

In this case indifference is calculated by:

$$[1 - \pi(SP)]*[e^1 + SP] + \pi(SP)*e^o = \int_0^{SP} pdf(P:SP)*[e^1 + P - e^o]\ DP + e^o$$

Collecting terms and multiplying through by minus 1,

$$e^0 - e^1 - SP_c = \frac{\int_0^{SP} pdf(P{:}SP)\,[e^o - e^1 - P]\;DP}{[1 - \pi(SP_c)]}.$$

or $$WTP - SP_c = \int_0^{SP} pdf\,[WTP - P]\,DP\,/\,[1 - \pi] \;\geq\; 0.$$

and $$\Delta CS \geq 0.$$

Note, $$WTP - P < e^o \text{ whenever } SP < e^o - e^1.$$

Therefore if $$\int_0^{SP} pdf\,[WTP - P]\,DP > 0, \text{ strictly,}$$

then $$WTP - SP_c > 0$$

Or $$SP_c < WTP \text{ whenever } \int_0^{SP} pdf\,DP > 0.$$

So when SP = WTP, the respondent is ***not*** indifferent. Rather it is a dominant strategy to reject the referendum. Indifference occurs at some referendum price (SP) strictly less than WTP. The introduction of a conservative response behavior affects the DC responses, but in a predictable fashion. It remains a dominant strategy to Reject whenever $SP > WTP_{true}$. To summarize the optimal response strategy:

$$\text{Reject if } SP > WTP,$$

$$\text{Reject if } SP > SP_c,$$

$$\text{Accept if } SP < SP_c,$$

be $$\text{Indifferent if } SP = SP_c.$$

where $0 \leq SP_c \leq WTP$.

For agents (perhaps only a few) that face referendum prices in the range $SP_c \leq SP < WTP$, it is optimal to record a conservative rejection of the referendum price even though the agent is willing to pay SP. Therefore, DC estimates of WTP will be conservative underestimates of WTP. A properly specified DC estimate will equal SP_c, not WTP. Yet if we want to identify SP_c and WTP or even to diagnose the presence of these voting strategies in a referendum, more information is required. De-briefing (or exit polls) to detect protest behavior from one shot dichotomous choice responses would identify only those agents who receive referenda prices falling above their critical level of indifference, SP_c, and below their true value, WTP.

CONTINUOUS RESPONSE DATA

If strategic voting operates on an actual vote, a WTP survey might reveal the same pattern of response. However diagnosing the existence of such strategic responses in either actual elections or in surveys is difficult with only dichotomous referenda data. Fortunately, unlike referenda, surveys provide an opportunity to follow up with voters. Exit poll surveys in actual elections can solicit continuous responses following a dichotomous choice decision. Similarly if the strategy found in public choice literature exists in dichotomous choice survey data, that strategy should influence open-ended responses as well. If so, analysts might be able to utilize open-ended responses to diagnose the presence of the voting strategy and, if theoretically consistent, even to infer valid WTP estimates.

Consider the expected payoffs for the open-ended question. The open-ended response offers an additional opportunity to improve expected payoffs. Facing a budget maximizing agency, voters know that higher Bids increase the likelihood that a program will be implemented successfully or, in the case of initial rejection, increase the chance of favorable revision. Yet by offering the public agency a larger budget, higher Bids also decrease expected benefits attributable to any eventual policy implementations that actually occur. In deciding how to answer an open-ended WTP request that immediately follows a dichotomous referendum question, the respondent needs to confront these competing effects in determining an optimal Bid.

Following YES Vote: The respondent faces the problem:

B.0 Min Expenditure(Bid) $= [1 - \pi(\text{Bid},\text{SP})]*[e^1 + \text{Bid}] + \pi(\text{Bid},\text{SP}) * e^o$

subject to: Expenditure $[\text{Bid}^*,\pi^*(\text{Bid: SP})] \leq$ Expenditure $[\text{YES(SP)}]$.

This expression outlines the competing effects on expected payoffs of Bids that follow SP acceptance. First of all, we know for SPs that lie below SP_c voters expect to realize real surplus gains. Furthermore these gains are strictly greater than payoffs expected from a referendum rejection to the same $SP < SP_c$. If agents who initially accept an SP can affect the chance that the policy eventually will be implemented, that change would mark a clear net benefit to those agents. B.O records this partial effect. Since by definition, SP < WTP means that $e^1 + SP < e^o$, a decrease in π reduces expected expenditures. $\pi(Bid;SP)$ in B.0 is reduced by ever higher Bids, meaning the probability of implementation increases - a benefit. Yet B.O also records the costs of achieving this impact on the approval success.

This change in $\pi(Bid;SP)$ is propelled by the offer of a larger budget - higher Bids. This means that the favorable influence of higher Bids on the policy process is not costless. Bids raise the expected costs of projects actually undertaken, expressed in B.O as an increase in e^1 + Bid. Optimal Bids unaffected by the constraint balance these competing influences such that :

$$\frac{\partial\, Ex}{\partial\, Bid} = 0 = [1 - \pi(Bid,SP)] - \frac{\partial\pi(Bid,SP)}{\partial\ Bid} * (e^1 + Bid - e^o).$$

Or

B.1
$$(WTP - Bid) = - \frac{[1 - \pi(Bid,SP)]}{\partial\pi(Bid,SP)\,/\partial Bid} > 0.$$

Since $\partial\pi/\partial Bid < 0$ & $[1 - \pi] > 0$, then B.1 > 0. This means the optimal Bid lies below WTP (i.e. $\Delta CS > 0$). Optimal Bids retain some strict surplus and are not expected to rise all the way to WTP.

More importantly, optimal Bids likely increase expected surplus outcomes above gains that voters already expected to realize by accepting the initial referendum.

The constraining condition on B.O is expected to bind only where $SP = SP_c$. The constraint itself simply requires that expected expenditures (Expenditure [Bid*,π*(Bid: SP)]) will not increase expenditures above the level secured by the original referendum (Expenditure [YES(SP)]) as a result of the Bid. At all other SP positions below SP_c, some benefits to bidding above SP are anticipated; and B.1 holds as an interior solution.

That agents associate the open-ended question with an opportunity to secure gains beyond those expected from the referendum question reveals an important feature of voters responding to an agenda setting agency. Under this hypothesis, voters do not view referenda prices as having been just plucked out of the air. Rather voters motivated by strategic concerns attribute the initial SP to an attempt by the agency to secure a generous budget. If the agency carefully selected that SP to best evidence a broad support for its initiatives at a reasonable budget, that same policy decision process will likely be strongly influenced by the first voluntary Bid movements above SP. Those initial contributions are especially helpful to the agency to evidence legitimate support for its proposed project.

It is also important to realize that the optimal surplus maximizing Bid is not constant across initial price suggestions $0 < SP < SP_c$. π, after all, is conditioned on SP. The influence of this initial signal continues to impact the optimal Bid that a respondent ultimately adopts. As agents embark on calculating their trade-offs between surplus (WTP - Bid) and the chance of approval, the initial SP will influence the optimal Bid position finally adopted. B.2 tracks this influence by:

B.2 $$\frac{d[WTP - Bid]}{dSP} = \frac{(\partial\pi^*/\partial Bid)^*(\partial\pi^*/\partial SP) + (1-\pi)[\partial(\partial\pi^*/\partial Bid)/\partial SP]}{[(\partial\pi/\partial Bid)]^2} < 0.$$

$\partial\pi^*/\partial SP > 0$ and $\partial(\partial\pi^*/\partial Bid)/\partial SP < 0$. $\partial\pi^*/\partial Bid < 0$. B.2 says that optimal responses begin to close the distance between WTP and Bid over SP, meaning Bids move closer to WTP if they first adjust from higher acceptable SPs.

Tracking the probability of policy success $(1-\pi)$ at *optimal* Bids across the range of suggested prices underscores an important feature of the hypothesized surplus maximizing voter strategy.

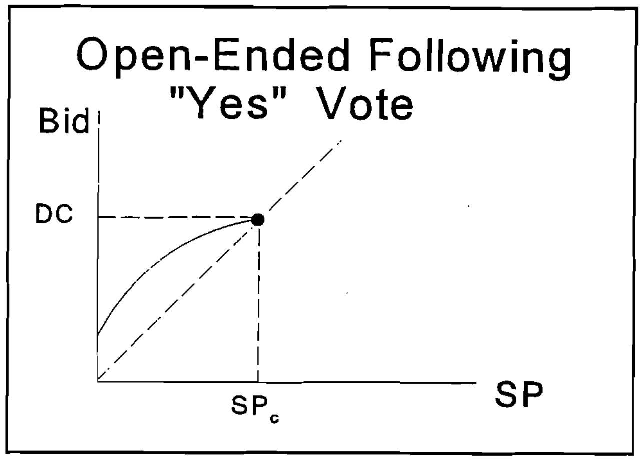
Respondents assess that the overall improvement realized from the entire sur- vey experience will be greater if those respondents receive smaller suggested prices initially.

Voters reacting to an agenda setting agency interpret low referendum prices as signals that low Bids near SP can still powerfully influence the overall chance of approval. This means voters attribute a greater overall impact to a Bid of equal magnitude which first starts from a lower SP. Since voters believe that agencies prefer to start with an SP as large as they believe can be favorably received and plausibly supported, respondents receiving smaller suggested prices feel free to offer Bids well below WTP that, nonetheless, still exert a powerful influence on the overall chance of project provision. Simply, with fewer perceived constraints on their dual objectives of reducing outlays, e^{j} + Bid, and affecting project success, 1 - π(Bid,SP), agents commencing from smaller SPs anticipate greater successes on both counts: they can affect a greater overall chance of approval with Bids that are still strictly smaller than Bids adjusting from larger SPs.

In B.2 $\partial\pi^*/\partial SP > 0$ emerges if voters maintain several assumptions consistent with responding to a budget maximizing agency. Respondents attribute a higher chance of success to an identical Bid if that Bid first adjusts from a lower SP. For example, a \$40 Bid is, overall, more persuasive if it adjusts from an initial price of \$20 than a \$ 40 Bid that followed acceptance of \$40 and never adjusted. Conversely, marginal movements from the same Bid level (eg. \$40) exhibit *marginally* greater impact on π if moving from a higher SP than from a smaller SP. That is, movements from SP=40 to Bid=41 marginally reduce π more than a one unit Bid increase from \$40 to \$41 that commences from SP=\$20. Hence, $\partial(\partial\pi^*/\partial Bid)/\partial SP < 0$.

Optimal Bids obeying this behavior would reveal a pattern in the data. Agents expect that ever larger surplus sacrifices are needed to lock in their initial gains from a referendum response as that initial referendum price request grows larger. So expenditure optimality in B.1 is characterized by a falling numerator, [1 - π(BID,SP)], and an increasing denominator [($-\partial\pi$(BID,SP)/∂Bid] as SP rises.

This implies a particular pattern of response following an SP acceptance.



Implication 1: Over the range of SP Acceptance from $0 \le SP \le SP_c$, Open-ended responses adjust upward from SP. The total magnitude of adjustments eventually fall as SP increases in the neighborhood of SP_c. Adjustments fall to zero where the suggested price equals SP_c.

This means that open-ended Bids will exhibit Starting point effects of a particular kind, illustrated on the adjoining figure. Bids rise over SPs, but at a non-increasing rate as SPs near SP_c. Bids eventually equal SP at the Accept / Reject indifference point (SP_c) where a DC estimator also identifies the estimated mean value of the latent dependent variable.

Following NO Vote: The optimal Bid response that follows a rejection of suggested prices which are higher than SP_c reduces to the problem:

Min Expenditure(BID) =

1 $\int^{SP} \{pdf(BID{:}SP)*[e^1 + P]\}\ DBid + (e^1 + Bid)\ [Pr(P=Bid)]$

2 $+\ \ e^o - \int^{SP} \{pdf(BID{:}SP)\ e^o\ DBid\} - e^o\ [Pr(P=Bid)].$

The probabilities sum to unity. Line one represents the range of expenditure outcomes possible if the issue is revisited after a NO vote. Line 2, of course, identifies the probability that the baseline expenditure is returned, meaning the NO vote and subsequent Bid permanently defeat the issue. In the referendum, the pdf carried SP as an argument. Here the pdf also carries Bid as an argument. So following a NO vote, the open-ended response can alter the cumulative probability that the issue will be revisited. Following a rejected SP, the chances of revisitation are positively related to Bids that position closer to SP than to zero. So,

$$\partial \, [\int^{SP} pdf(BID{:}SP)] \, / \, \partial \, Bid > 0.$$

Bids not only affect the entire likelihood of policy revisions; but they shift the distribution of outcomes. The pdf logically does not increase or decrease symmetrically as Bids rise and fall. Rather Bids shift the pdf toward a new mode, P = Bid.[3] So higher bidding to benefit an improved re-issuance of the proposal once again comes at a cost. The likelihood of any resulting cost revisions become centered more densely on the Bid offered to the agency. That is, higher Bids favorably impact the overall chance of project revision; but the budget-maximizing agency is more likely to take the respondent up on the higher Bid offered than to suggest alternatives even cheaper than that Bid. So $\partial Pr(P{=}Bid) / \partial Bid > 0$ as well. Once again increasing the probability that an issue will be reconsidered necessitates more expected surplus to be sacrificed to affect that outcome.

The optimal Bid minimizes expected expenditures such that:

$$0 = \int^{SP} pdf'_{Bid} (Bid{:}SP)*[e^1 + P] \, DBid - \int^{SP} pdf'_{Bid} (BID{:}SP)* e^0 \, DBid$$
$$+ Pr'_{Bid}(P{=}Bid)[e^1 + Bid] - Pr'_{Bid}(P{=}Bid) \, e^0 - Pr(P{=}Bid).$$

Collecting terms, this optimality condition outlines the trade-off as:

$$\int^{SP} pdf'_{Bid} [P - WTP] \, DBid \; - \; Pr'_{Bid}(P{=}Bid) \, [WTP - Bid] \; + \; Pr(P{=}Bid) = 0.$$

Or the marginal benefits of increasing Bids to increase the overall probability of expenditure declines ($\int^{SP} pdf'_{Bid} [P - WTP] \, DBid$) are exactly offset by the marginal sacrifices to surplus if those improvements actually occur ($Pr'_{Bid}(P{=}Bid) [WTP - Bid]$). Greater chance of gain is realized by lowering the magnitude of gains; or marginal benefits equal marginal costs at the optimal Bid. Re-writing, we define:

B.3 $$\int^{SP} pdf'_{Bid} [(e^1 + P) - e^0] \, DBid \; = \; Pr'_{Bid}(P{=}Bid) \, [WTP - Bid] \; - \; Pr(P{=}Bid).$$

3. Without loss of generality for our purposes, this effect is captured as a discrete probability that the revisited price (P) will equal the Bid. In a continuous distribution, the exact point probability that P=Bid will of course be zero. The probability modelled above intends only to illustrate the increased density of the pdf on Bid once an open-ended response is offered. Since the researcher doesn't need to observe directly the respondents' expectations either to infer WTP or to diagnose the presence of this behavior, the simplification is extended without loss of generality.

Bids are optimized where the marginal expenditure effects from a change in the probability (∂ pdf / ∂Bid) of a successful proposal revision [(P-WTP) or (e^1+P-e^0)] are offset by a change in expected consumer surplus (WTP - Bid) of that successful revision. As long as outcomes on the LHS of B.3 express real surplus gains (P - WTP $<$ 0), the optimization property above motivates Bids to keep rising through SP. This implies through the range $SP_c < SP < WTP$ Bids are non-declining and continue to rise.

Following SP rejection in the range $SP_c < SP < WTP$, votes consistently adjust downward. Nonetheless the measure of final Bids does increase through SP until SP=WTP. Yet Bids don't rise forever.

SP will eventually lie above WTP. At this point, some of the possible revised options [P - WTP] under the pdf constitute real injuries. Following a No vote to an SP > WTP, it is important that some expected reconsidered prices now include expenditure increases [P - WTP > 0; or (e^1 + P) - e^0 > 0 in B.3]. The value of these expected losses are measured as $\int_{SP}^{WTP}$ pdf [P - WTP] DBid. As SP exceeds WTP, agents no longer maximize surplus with ever higher Bids; but become increasingly concerned to avoid real losses from baseline utility, e^o. Beyond SP > WTP, lower Bids help reduce the probability of losses by diminishing both revision prospects overall and by moving the modal center further from SP toward zero - out of the range of WTP $<$ P $\leq$ SP. Technically the expected expenditure reduction benefits of high Bids in the LHS of equation B.3 begin to fall after SP=WTP. Therefore optimal Bids also begin to fall unless the entire pdf happens to take on all realization exactly at Bid. The prospect that a rejected proposal could re-appear with a price tag above WTP induces, more and more, a desire to defeat that proposal permanently. Simply as SP extends beyond WTP more and more, revisions themselves carry the increasing possibility of real welfare losses - losses that continue to grow in magnitude as the SP increases. This implies a general shape to the Bid function.

Implication 2: Over the range of SP Rejection, Bids increase through SP to a peak at SP = WTP. Following SP = WTP, Bids do not rise further and may decline.

SURPLUS MAXIMIZING ADJUSTMENTS

The two implications of this behavior mark a very specific trail in the data that can reveal WTP inferences. If specific internal consistency checks between DC responses and open-ended responses prevail, this surplus preserving behavior implies:

PROPOSITION 1: Bids influenced by the surplus preserving behavior identify expected Bid = SP at a point not different than a correctly specified DC estimate.

The intersection identifies SP_c. A relation between BID and SP influenced by surplus preserving strategies will predict that open ended responses will cross the 45 degree line in BID/SP space at a point not statistically different from the mean limited dependent variable estimator using the preceding DC responses. Further,

PROPOSITION 2: If respondents exhibit Surplus Maximizing Behavior in the Open-Ended array, then the distance between the peak of the BID function and the 45 degree line in Bid/SP space records the magnitude of conservative referendum responses. A valid inference of WTP is equal to the SP at the point Bid reaches its maximum.

Other behaviors can explain an increasing, then decreasing Bid/SP relation [26, 28]. Yet we can infer WTP associated with the maximum of this relation only under specific conditions. Open-ended and referendum responses must be consistent. Bids also increase through SP in the neighborhood below the DC estimate. Moreover, open-ended responses equal SP within a statistically significant confidence interval from the DC mean estimate. Even further, surplus maximizing strategies require that the peak Bid of the Bid/SP relation in Bid/SP space occurs at an SP strictly greater than or equal to the mean of the DC estimate.[4]

Only after meeting the particular conditions of this filter can we infer WTP equal to the SP corresponding to the estimated peak of open-ended responses over the range of SP.

Propositions 1 and 2 above allow several general shapes for open-ended data consistent with surplus maximizing adjustments from SP. The cases depend on the attributions that the SP is a take-it-or-leave-it proposition in the DC response and the probabilities attributed to the decisiveness of this exercise as higher open-ended responses are given.

CASE ONE: Satisfactory Referenda, 'Asymptotic' Bids. This strategy is illustrated on Figure 1. This strategy implies a wholly satisfactory DC response. Following the DC response, respondents attempt to preserve expected gains from YES voting and increase the likelihood of obtaining the service by increasing Bids until WTP is realized. Then respondents begin to reject referenda prices and follow-up Bids flatten out over the remaining SP range. Simply the aggressive proposal defeating behavior is absent from both data array. The Bid peak is realized at a nondifferentable point where the predicted Bid also intersects SP. Further this asymptote/intersection will occur at a point not different from a properly specified WTP estimate from the DC array. These properties are still consistent with both propositions. The Bid = SP intersection is the DC estimate. In this case, the intersection is also the Bid peak and identifies a WTP inference.

It is assumed that such a discontinuous function can be approximated by a non-linear asymptotic function which is continuous in BID/SP space.

4. The BID function must also pass a test for "dominant" SP anchoring. That test is illustrated in the example herein.

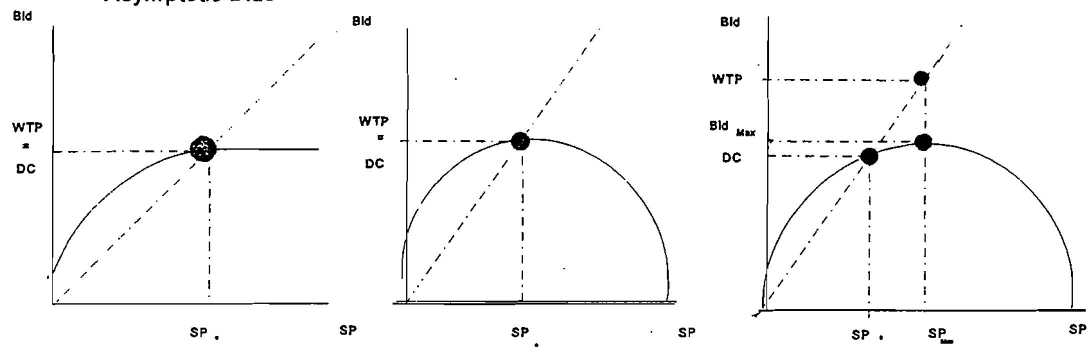
Surplus Maximizing Response Behavior

Three Cases

Figure 1. Satisfactory Referendum, Asymptotic Bids

Figure 2. Satisfactory Referendum, Conservative Bids

Figure 3. Conservative Referendum, Conservative Bids



CASE TWO: Satisfactory Referenda, Conservative Bids. It is plausible that open-ended bids are affected by the conservative surplus maximizing behavior only with the introduction of the follow-up questions. In this case the Bid peak will occur at a point not statistically different from the Bid = SP intersection. The conservative proposal defeating behavior *only* emerges with the continuous Bid opportunity. Though take-it-or-leave-it motives dominate the DC array; conservative Bids induce the Bid function to begin to fall after SP = WTP (Proposition 2) as in Figure 2. This mitigates possibilities of re-visited proposals at real costs in excess of WTP. The distance between the Bid peak and the Bid = SP intersection collapses to zero. So satisfactory referenda assures that the peak occurs at the Bid = SP intersection which is still the DC estimate (Proposition 1).

CASE THREE: Conservative Referenda, Conservative Bids. This case includes all aspects of the voting behavior consistent with the public choice literature on this subject. This case permits conservative behavior in both the DC array and the Bid array. Respondents begin to reject referendum prices below WTP and adjust downward from SP. The intersection of predicted Bid and SP occurs at Bids strictly less than WTP as in Figure 3. This forces the peak to the right of that SP position (proposition 2).

The Bid = SP intersection conforms to the DC estimate of the mean of the latent dependent variable as in Proposition 1; but that point no longer represents a proper WTP inference. Bids no longer peak at SP_c . They continue to rise until SP = WTP. The estimate of WTP still occurs where the value of SP where Bids reach their maximum (Proposition 2); yet the distance between WTP and SP_c is positive. So the distance between SP_{max} and SP_c records the magnitude of an underestimate of WTP retrieved from the DC array.

EMPIRICAL ILLUSTRATION

In all three cases the open-ended peak constitutes a valid WTP inference. In all three cases the Bid = DC intersection also equals the DC estimate. Discerning the case distinctions requires something beyond the DC responses; namely, the Bid array.

There are three steps to an analysis of the continuous Bids to diagnose this voting behavior: 1) specifying the Bid/SP relation; 2) testing if Bid variation is dominated by anchoring; and, finally 3) inferring WTP. The remainder of the paper illustrates these steps.

Used to illustrate these diagnostics is a contingent value study of recreation users of an urban park [27]. Users were intercepted at fishing sites, hiking trails and picnic areas over one summer. A mail survey of these users followed in the Autumn. The survey asked respondents to value several proposed improvements in various attributes of the park in a discrete-continuous framework. See appendix 1 for survey details.

1) Specifying the Bid/SP relation

Specification of the Bid/SP relation must distinguish among the cases presented. At the same time it must distinguish itself from anchoring effects per se.

Testing various polynomial forms for the Bid/SP relation, LM tests recorded on Table I strongly suggest SP effects are significant in the data. The linear SP form, generally consistent with models of anchoring that predict Bids continue to rise, retains some explanatory power; but a 4th order SP polynomial performs better. An LM test for the three restrictions (P^2, P^3 and P^4 equal to zero) yields a $\chi 2_{df=3} = 14.27$[5] value for difference in explanatory power over the linear model.

This suggests that the data rises, reaches a peak and then falls. The estimated 4th order polynomial predicts that Bids peak at SP = \$ 93 and then move toward zero. This weakly supports cases 2&3 rather than the asymptote (case 1). Case 1 is suggested by a polynomial that moves toward $+\infty$, $-\infty$, or zero beyond the range of reasonable mean and median WTP inferences (flattening over the observed SP range). That array may be asymptotic. The peak Bid (close to \$ 53.50 where SP = \$ 93) is clearly within the magnitude range of the logit estimate (\$46.80), the median (\$ @ 80), and observed range of SPs (which includes \$75 and \$150). This is clearly a weak test.

A more direct test of case 1 embeds all three cases into a single functional form. That form is flexible enough to test many shapes consistent with theories about the Bid/SP relation, including anchoring effects. Consider the following adjustment as one suggested example:

NL. $$\text{Bid} = [K * SP / (SP + A)^{\rho}] + C.$$

ρ is the parameter of interest. It indicates the general shape of the relation between Bid and SP. An asymptotic form implies $\rho = 1$. Cases 2 and 3 imply $\rho > 1$. For completeness, Bids anchored on SP are expected to increase inside a reasonable range of WTP.

---

5. Under the assumptions generating the confidence intervals for the χ^2, the critical value for $\chi^2_{df=3,0.05} = 7.815$.

TABLE I

Specifying the Suggested Price Effect

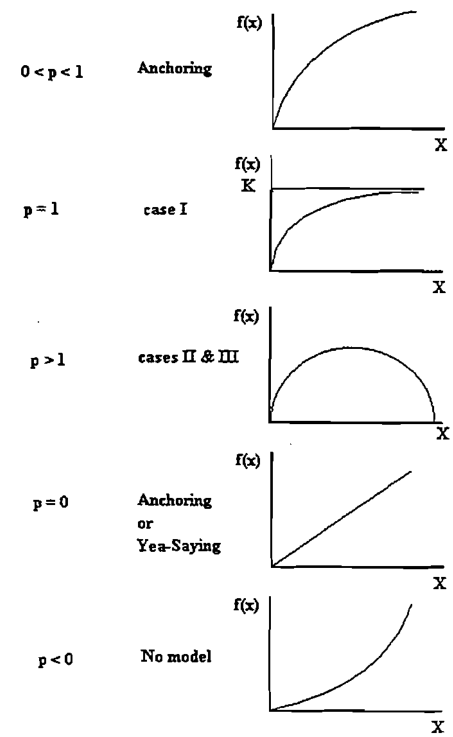
Model	No SP Effect	Linear SP Effect	4^{th} order SP Effect
N[a]	274	274	274
R^2_{adj} (%)	16.52 %	26.53 %	29.56 %
SSE[b]	$3.4609*10^6$	$3.0348*10^6$	$2.8767*10^6$
LLF[c]	-1367.15	-1349.15	-1341.82
Variable Name	Estimated Coefficient (Standard Error)		
Constant	-0.54 (11.29)	-16.51 (10.91)	-33.95 (30.16)
Income	1.86 (1.054)	1.59 (0.989)	1.68 0.97
Family Size	-3.31 (1.64)	-3.20 (1.54)	-2.62 (1.53)
Education	8.21 (1.76)	8.05 (1.65)	8.14 (1.63)
Age	0.319 (0.667)	0.418 (0.626)	0.194 (0.630)
Age^2	$-8.57\ 10^{-3}$ $(7.86\ 10^{-3})$	$-8.94\ 10^{-3}$ $(7.38\ 10^{-3})$	$5.68\ 10^{-3}$ $(7.44\ 10^{-3})$
Sex	-8.738 (4.43)	-7.416 (4.16)	-6.748 (4.09)
Price	-	0.262 $(4.30\ 10^{-2})$	1.164 (4.10)
$Price^2$	-	-	$-6.63\ 10^{-3}$ (1.62)
$Price^3$	-	-	$-1.76\ 10^{-5}$ $(2.10\ 10^{-3})$
$Price^4$	-	-	$1.50\ 10^{-7}$ $(7.99\ 10^{-6})$
Test	$Price=Price^2=Price^3=Price^4=0$	$Price^2=Price^3=Price^4=0$	
LM Stats (X^2)	46.25	14.27	

[a] Trimmed Top 5% Bids
[b] Sum of Squared Errors
[c] Log of the Likelihood Function

This implies $\rho < 1$. Various forms linking Bid shapes to ρ values are illustrated on figure 4.[6] The non-linear regression estimate of ρ equals 2.7197 with a standard deviation of 0.1145. A simple t-statistic for the distance from $\rho = 1$ is 15.02, suggesting the asymptotic form unlikely characterizes this distribution. The distance from zero (linear form) finds a t-value of 23.75. Strongly anchored behaviors are generally consistent with values of ρ between zero and unity.

Using both tests, we conclude that the Bid/SP relation is marked by a general shape that increases through very low SPs and then decreases for large SPs. Results are outlined in Appendix 2. Given the well-known problems of non-linear in parameter regressions, subsequent investigations are conducted on the polynomial form.

Figure 4.



2) Testing information in BID variation

Adopting the higher order polynomial specification of the Bid/SP relation, we investigate if Bid variation is statistically different from an array dominated by anchoring effects. The specification results favoring the higher order polynomial form do not erase the prospect that residual anchoring, more consistent with an ever increasing functional form, might still affect the Bid/SP relation so severely as to obviate uses of open-ended data to reveal response strategies. For completeness we test this possibility as well before drawing WTP inferences.

6. This form also detects various degrees of anchoring. Bids that reflect adjustments from SP toward WTP but are significantly anchored on SP are expected to continuously increase inside the neighborhood of WTP [24]. Anchoring theory predicts $0 <= \rho < 1$ and $A > 0$. Some recent results in cognitive psychology [27] suggest anchoring can allow bids to eventually decline; but this decline occurs only when SP values are considered so extreme and so unreasonable to the respondent (i.e. well beyond WTP) that the SP no longer carries information to the respondent and consequently no longer anchors a response.

Dominant anchoring predicts that BID = SP will occur at *median* DC acceptance, rather than *mean* DC acceptance [10]. Purely random Bid adjustment from SP means the open-ended array tracks only anchoring effects, not behavior tracked by the DC estimator. So we test if the estimated value of Bid at the Bid/SP intersection differs from the DC median.

The polynomial estimated intersects the 45 degree line at approximately 45.5. Inspection shows 54.55 % of the respondents still accept $ 75. The difference between 45.5 and a median of $ 80 generates a 79% confidence in the difference (t=0.804) under assumptions of the t-statistic. The test uses a confidence interval for a predicted point (Bid = 80) outlined by Kmenta [18]. See endnote[1]. If the median is higher, more confidence is implied. Added confidence is limited as only one SP between $ 30 and $ 150 exists in the data set. Results should be interpreted as illustrative of this diagnostic process for dominant anchoring.

WTP INFERENCES:

The estimated Bid function reaches a maximum where SP ≈ 93. This is the implied WTP value for the Surplus Maximizing behavior consistent with case three: Conservative Referenda, Conservative Bids. Proposition one predicts Surplus preserving behaviors will identify predicted Bid = SP equal to an appropriately specified DC estimate. Expected Bid = SP at $ 45.5. The DC logit estimate by Wu estimates WTP = $ 46.8. Proposition two demands that Bid peak ($56.5) be greater than or equal to both the intersection ($ 45.5) and the logit estimate ($ 46.8) and that all three will be less than the inferred SMA estimate ($93), meaning the peak of the bid function will not be to the left of the 45 degree line.

The values ($ 56.5 and $ 46.8) are statistically indistinguishable (t-statistic ≈ 0.226 for difference), meaning Cases 2 and 3 are indistinguishable. Case 2 is consistent with WTP between $ 45.50 and $56.50 (or the discrete choice estimate of $ 46.80). With the only suggested price between $ 30 and $ 150 equal to $ 75, reasonable confidence in precise WTP inferences and detection of statistical difference between $ 45.5 and $ 56.5 is unrealistic. Our main objective is to test response consistency with DC estimates predicted by strategic voting. The tests fail to reject consistency and suggest conservative Bid responses to suggested prices.

For completeness WTP inferences for other models are listed in Table 2.

Table 2

Inferences to Willingness to Pay

Estimated Model	Inferred Estimate
$WTP_{discrete}{}^{a}$	\$ 46.80
$WTP_{case\ 3}$	\$ 93.00
$WTP_{case\ 1}{}^{b}$	\$ 56.50
$WTP_{anchored}{}^{c}$	\$ 45.50

a Estimated in Wu [27].
b Approximate peak using predicted coefficients at SP = 45.5.
c Following Thayer [25].

In this illustration, results are consistent with the voting strategy suggested. The SMA model better predicts the SP effect specification and is considered the most preferred model for these test results. Yet the paucity of SPs between \$ 30 and \$ 150 (only \$ 75) explicitly limits confidence in the particular \$ 93 point estimate.

This illustration demonstrates a fairly strict set of conditions to which CVM data would have to conform to plausibly infer the Bid adjustment behaviors suggested. Beyond simply predicting shape, there are stronger empirical implications that we want to re-emphasize. Our three surplus preserving and maximizing voting models predict a rigid conformity between DC and Bid responses. First, the asymptotic model predicts: 1) an asymptote to the Bid Function, 2) an asymptote value not different from a correctly specified DC estimate, and 3) an estimated Bid = SP not different from either the asymptote or the DC estimate. SMA behavior predicts: 1) the Bid function will generate a discernable peak (maximum), 2) the Bid=SP intersection is not different from a correctly specified DC estimate, and 3) the peak will not occur "to the left" of the point where the predicted Bid equals SP (the intersection). Finally, in all cases, the data must identify a Bid = SP intersection that is statistically distinguishable from the predicted median acceptance level in the DC data. These, we argue, are compelling standards derived from voting theory.

CONCLUSION

Starting price effects in contingent value data sets are frequently considered an irreconcilable bias. The assumption that SP effects emerge solely from cognitive response anchoring often motivates a deference to referenda data. Further open-ended SP affected responses are seen to cause systematic overestimates of WTP. Rejecting anchoring as the sole source of strong SP effects, we challenge both conclusions.

The economically induced SP effect models suggest that it is unreasonable for respondents who take seriously the policy implications of a CV survey to voluntarily exhaust their entire residual surplus by offering Bids all the way to WTP. Rather some conservative bidding seems possible. The level of that underbidding varies with the SP signal. This produces an economically motivated SP effect. Three models are developed for this phenomenon. Each leaves a different and distinct empirical trace - distinguishable not only from each other but from pure anchored SP effects generally. Surveys whose proposal or payment vehicle are incredible to the respondent would be perceived as largely hypothetical and would unlikely exhibit the data patterns outlined in this work.

The implications for double bounded surveys formats is not addressed; but the voting strategy considered herein is consistent with discoveries that double-bounded questionnaires are from different distributions. This discussion suggests there may be promising theoretical insight to the question of reconciling two sets of DC responses along more these theoretically consistent lines.

There are numerous potential sources for starting point effects. Those introduced herein demand a rigid set of internal consistencies between initial referendum responses and continuous bid responses. If follow up responses are influenced by suggested prices through the behaviors introduced, these cross-consistencies constitute a powerful validity check on both DC and continuous data.

In testing for the voting strategy discussed, there are several important practical implications. Surveys have to be assessed individually; and many data sets may not conform to the behavior suggested. More importantly, there is an implied trade-off when choosing the range of SPs adopted in the survey design. Extreme SPs are often required for one shot DC estimation; yet those same extreme SPs reduce the plausibility that the survey will exhibit surplus maximizing voting behavior at those prices. The empirical illustration above demonstrates the difficulties in employing the diagnostics on open-ended data without an adequate number of suggested referenda prices in the range of the mean and median DC responses.

Further investigations using both continuous and referendum data together to improve M.L.E. specifications, obtain better SP specifications, and eventually to develop more efficient WTP inferences by employing *all* of the response information available in the data are areas for profitable exploration. It seems reasonable that certain payment vehicles or questionnaire formats are associated with different types of SP effect models. This demands an important role for more deliberate, targeted respondent de-briefing to follow up WTP questions. Beyond reviving the use of continuous data with its potential for more efficient and less expensive CVM experiments, analytic investigations into SP effects might teach us a lot about respondents' reactions to questionnaires by economists.

Appendix 1

Descriptive Statistics for Independent Variables

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
INCOME	274	3.9854	2.4057	5.7873	1.00	12.00
FZ	274	3.1058	1.5406	2.3734	1.00	11.00
ED	274	5.7920	1.5656	2.4511	1.00	8.00
YR	274	40.595	15.874	251.98	20.00	89.00
YR2	274	1899.0	1373.8	0.1887E+ 07	400.00	7921.0
SEX	274	0.489	0.5007	0.2508	0.00	1.00
W32	274	36.646	39.406	1552.8	0.00	150.00
PRICE	274	52.153	47.856	2290.2	10.00	150.00

Bid is the respondent's offer (w32) and SP is price. Income is an ordered variable from one to 12, ED is an ordered variable from one to eight recording level of education. Sex is a one / zero variable with one representing Male. Continuous IVS are: FZ , the size of family; and YR, the age of the respondent. Higher order price variables are labelled price, meaning price.

Response rate is 47.22 %. Non-respondents are known to be recreationists. Data also includes zip code. WU [27] reports no significant difference in location between respondents and non-respondents.

VALUATION QUESTION:

third question in embedding sequence. three goods added cumulatively.

"Would you vote to maintain SVS readings at 80, to improve stream beds visibility to 20 percent, **and** to add 3 miles of new hiking trails (see table at the bottom of this page) if it cost households like yours $ ___ each year in additional taxes ?

YES <u>For</u> the $ ___ permanent increase in my annual taxes needed to keep SVS readings around 80, to raise stream bed visibility to 20 percent, and to increase the hiking trail network from 5 to 8 miles.

<u>NO</u> <u>Against</u> the tax-financed pollution control and hiking trail expansion program.

What is the <u>maximum</u> permanent tax increase you would pay to maintain biological diversity, to improve stream bed visibility, and to increase hiking trail mileage ?

I would pay up to $ ___ in additional taxes each year so that SVS readings would stay at 80, instead of falling to 65, stream bed visibility would improve from 0 percent to 20 percent, and the hiking trail network would be extended from 5 miles to 8 miles."

Note: Set of Suggested Prices = {$10, $20, $30, $75, $150}

Appendix 2

NLP Regression Test

of Functional Form

TRIM TOP 5%

$$\text{Bid} = \frac{(K1*I + K2*YR + K3*YR^2 + K4*ed + K5*fz + K6*sex)*\ SP}{(SP+A)^P} + C1*I + C2*YR + C.$$

FINAL STATISTICS : Shazam NLP procedure

LOG-LIKELIHOOD FUNCTION= - 1382.007

	COEFFICIENT	ST. ERROR	T-RATIO
K1	0.96224	1.0000	0.96221
K2	0.89389	1.0000	0.89387
K3	-1.4953	0.38365	- 3.89760
K4	0.98113	1.0000	0.98113
K5	0.98021	1.0000	0.98020
K6	0.99768	1.0000	0.99768
A	3.3548	1.0558	3.17740
P	2.7197	0.11450	23.75300
C1	5.1752	0.77264	6.69810
C2	0.52793	0.12056	4.37900
C	2.3137	1.0218	2.26430

Initial values on "C" were chosen to reflect coefficient results in the 4^{th} order linear regression. P was also set at unity. Not surprisingly, variation in "K" (see NL. p.27) is dominated by "C" and the 'shape' parameters of the SP effect (A and P). The estimated value of C at the mean is \$ 44.37. Mean Bid is \$36.65.

The size of individual gradients along with the coefficient and standard error results suggest that parameter estimates are likely sensitive to initial value suggestions. Results are reported from a single attempt. Specific results herein are offered for largely illustrative purposes. Nonetheless the functional form introduced seems promising in its flexibility. The data set employed has so few starting prices and a wide dispersion among responses at the highest price suggestion that non-convergence seemed (seems) likely. Future explorations using this form may be helpful. One possibility, with some theoretical appeal, is to separate the SP effect shape variables entirely from interactions with other independent variables. That is, specify "C" consistent with regression coefficients in the DC estimate and estimate "K" as a single parameter constant. "C" then would act as an intercept to an asymptotic form.

REFERENCES

[1] Alberini, Anna. 1995. "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bounded, Double-bounded, and Bivariate Models." *Journal of Environmental Economics and Management*. 28:3. p.287 - 306.

[2] Arrow, Solow, Portney, Leamer, Radner and Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* v. 48 p. 4601 - 4614.

[3] Bendor, J.; Taylor, S., Van Gaalen, R. 1985. "Bureaucratic Expertise versus Legislative Authority: A Model of Deception and Monitoring in Budgeting," *American Political Science Review*. v. 79. p.1041-1060.

[4] Borcherding, T.E., Pommerehne, W.W., Schneider, F. 1982. "Comparing the Efficiency of Private and Public Production: The Evidence from Five Countries," *Zeitschrift fur Natioalokonomie*. v. 89. p. 127-156.

[5] Boyle, K.J., Bishop, R.C., and M.P. Welsh. 1985. "Starting Point Bias in Contingent Valuation Bidding Games". *Land Economics* 61: 188-196.

[6] Bishop, R.C., and T.A. Heberlein. 1979. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61:5 926 - 30.

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

[8] Cameron, T.A., and Quiggin, J. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-Up' Questionnaire, " *Journal of Environmental Economics and Management.* v 27. #3. p. 218 - 234. November, 1994.

[9] Carson, R., and Hanneman, M. "Use of Simulated Political Markets to Value Public Goods. " Working paper. Economics Department, Univ. of California - San Diego 1986.

[10] Eavey, C.L., and Miller, G.J. 1984. "Bureaucratic Agenda Control: Imposition or Bargaining?" *American Political Science Review*. 78: 719-33.

[11] Farmer, M.C., and Randall, A. 1992. "Starting Point Effects in CV Data Sets." discussion paper presented at the summer meetings of the *American Agricultural Economics Association*. Baltimore, MD.

[12] Filimon, R., 1982. "Asymmetric Information and Agenda Control." *Journal of Public Economics*. 17: 51-70.

[13] Fort, R.D. 1988. "The Median Voter, Setters, and Non-Repeated Construction Bond Issues," *Public Choice*. 56:3. 213-31.

[14] Hanneman, W.M. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics* 66: 332-41.

[15] Hanneman, W.M., Loomis, J., and Kanninen, B. "Statistical Efficiency of Double-Bounded dichotomous Choice Contingent Valuation," *American Journal of Agricultural Economics*. v. 73. p. 1255-1263. 1991.

[16] Herriges, J.A., and Shogren, J.F. "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning," *Journal of Environmental Economics and Management.* v. 30 # 1. p. 112 - 131. 1996.

[17] Hoehn, J.P., and A. Randall. 1987. "A Satisfactory Benefit Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management.* 14: 226-247.

[18] Kmenta, Jan. 1986. *Elements of Econometrics.* 2nd ed. New York. Macmillan Publishing Company.

[19] Mueller, Dennis. 1989. *Public Choice.* New York, N.Y.: Cambridge University Press.

[20] Romer, T., and Rosenthal, H. October 1979a. "The Elusive Median Voter," *Journal of Public Economics.* 12. p.143-170.

[21] Romer, T., and Rosenthal, H. November 1979b. "Bureaucrats versus Voters: On the Political Economy of Resource Allocation be Direct Democracy." *Quarterly Journal of Economics.* 93: 563-87.

[22] Romer, T., and Rosenthal, H. 1982. "Median Voters or Budget Maximizers: Evidence from School Expenditure Referenda." *Economic Inquiry.* 20: 556-78.

[23] Shapiro, P., and Sonstelie, J. 1982. "Representative Voter or Bureaucratic Manipulation: An Examination of Public Finances in California before and after Proposition 13." *Public Choice.* 39:1. 113-42.

[24] Slovic, P., and Lichtenstien, S. 1971. "Comparison of Bayesian and regression approaches to the study of information processing in judgement." *Organizational Behavior and Human Performance.* v. 44 p. 68 - 82.

[25] Thayer, M. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence," *Journal of Economics and Management.*

[26] Whitehead, J., Hoban, T.J., and Clifford, W.B. "Measurement Issues with Iterated, Continuous/Interval Contingent Valuation Data," Working Paper. East Carolina State Univ. Revised June 1994.

[27] Wu, Pei-Ing. 1991. *Benefit Estimation for Complex Policies: An Application to Ohio's Big Darby Creek.* Ph.D. Dissertation, The Ohio State University.

[28] Zellerman, O., and Fischbeck, P.S. 1995. "Over the Edge: A Quasi-Bayesian Model of Anchoring and Adjustment," working paper. Department of Social and Decision Sciences. Carnegie Mellon University.

ENDNOTE:

. To test the difference between the median and the estimated intersection of Bid and SP, predicted values for the independent variables at the estimated intersection were used. A median of $80 was also identified. Estimating Bid at the intersection follows Kmenta [17]. The VAR-COV matrix of regression "betas" identifies the appropriate coefficient values for the four orders of SP variables at the intersection (regression estimates of course are valid at the mean). As SPs were randomly distributed by survey design, the simple means and coefficient values of the other independent variables was used. Very low partial correlations on SP and other independent variables, incalculably small cov to price, confirms the sample randomness for SP in this assumption. Finally a confidence interval for the *forecasted* value of Bid at the P=45.5 intersection as different from the median $80 realizes at a t-value of approximated 0.80. Under normality assumptions this represents a 79% confidence in the difference.

ESTIMATING WTA USING THE METHOD OF PAIRED COMPARISON AND ITS RELATIONSHIP TO WTP ESTIMATED USING DICHOTOMOUS CHOICE CVM

John Loomis
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, CO 80523

George Peterson, Thomas Brown, Patricia Champ and Beatrice Lucero
Rocky Mountain Forest and Range Experiment Station
USDA Forest Service
Fort Collins, CO 80524

May 1996

ABSTRACT

The method of paired comparison is introduced as a means to estimate willingness to accept. The technique is adapted from psychology and involves having the individual make choices between two alternative gains. To estimate monetary willingness to accept (WTA) the individual is asked whether he or she would prefer to receive $X or a particular good. This question is repeated at several dollar amounts and the multiple bounded logit model is used to estimate WTA. In this study, the method of paired comparison is performed to estimate hypothetical WTA for an art print. The estimated mean (median) WTA is $59($52) using a parametric estimator and $66($34) for mean (median) using a non-parametric estimator. Applying a standard single bound dichotomous choice contingent valuation method to estimate willingness to pay (WTP) yields mean (median) of $28(28). Confidence intervals around the respective paired comparison and dichotomous choice estimates indicate these are significantly different. Nonetheless, the ratio of WTA to WTP is less than has been found in most past CVM studies and closer to ratios found when real money changes hands. The relative divergence of hypothetical WTA from actual cash WTP estimated in a separate treatment ($11), is also smaller using the method of paired comparison than in past dichotomous choice contingent valuation studies. While our study is exploratory in nature, it appears that the method of paired comparison warrants further research as a promising method to elicit WTA.

THE ISSUE: OBTAINING ESTIMATES OF WTA

When non-marketed natural resources legally owned by the public and managed by the government on the public's behalf are damaged, willingness to accept (WTA) is usually the appropriate measure of loss. Many economists believe that willingness to pay (WTP) will be a good proxy for WTA for goods, the benefits of which make up a small percentage of income or which have ample substitutes (Willig, 1976; Randall and Stoll, 1980). Hanemann (1991) suggests that when either of these conditions are not met WTA would be expected to exceed WTP (Hanemann, 1991).

Experiments designed to test the relationship between WTA and WTP have found estimates of WTA to be 2-10 times larger than WTP even with "trivial" market goods such as coffee mugs which make up a small part of income and have numerous substitutes. The divergence between WTA and WTP persist even with actual cash experiments (Welsh, 1986) and when goods are actually exchanged (Knetsch and Sinden, 1984). Kahneman, et al. (1990) suggests psychology provides explanations for these large disparities in the form of endowment effects and loss aversion. Horowitz, McConnell and Quiggin (1996) develop hypotheses and experiments to test between psychological explanations and those deduced from economic theory. These authors find only limited support for economic or psychological explanations. Another interpretation of the divergence is that it arises from the experimental design which usually involves asking individuals not in the market for the good and the WTA group receiving the good for free (Lucero, 1996). At this time, the consensus remains (Mitchell and Carson, 1989; Arrow, et al., 1993) that current forms of contingent valuation (CVM) do not appear capable of reliably measuring WTA.

In this paper we use the **method of paired comparison** (PC) to estimate WTA from a chooser reference point. By chooser reference point we mean the individual makes a binary choice between two alternative gains. For example, the individual would choose whether to accept $50 or an additional quantity of a public good. If he or she selects the public good, then WTA for the public good is inferred to be greater than $50. The chooser reference point avoids apparent loss aversion associated with the standard

CVM approach for measuring WTA as the individual is trading off two alternative gains. It may also be possible to value uncompensated losses using PC by asking whether the individuals would prefer $50 or that five acres of damaged wetlands be restored.

The purpose of this paper is to compare WTA elicited using the method of paired comparison to WTP elicited using a standard dichotomous choice CVM approach. WTA in a hypothetical market for an art print is compared to WTP in a hypothetical market and WTP in a real cash market. We are interested in whether the ratio of WTA (elicited by the method of paired comparison) to WTP (elicited by dichotomous choice CVM) is closer than past studies. This analysis is exploratory and designed to determine whether further research on the method of paired comparison is worthwhile.

USING PAIRED COMPARISON TO MEASURE WTA

Figure 1 illustrates the welfare constructs associated with WTP and PC WTA. The individual starts at Point A with two units of the private good (e.g., the numeraire) and four units of the public good (e.g., acres of open space). Then the individual is offered either three more units of the numeraire or three more units of the public good. If the numeraire good is chosen over the increment of the public good, the individual's WTA to forego the gain of AB units of the public good is greater than the value of AB units of the public good. In Figure 1, AC is the minimum amount of the numeraire good that would provide the same level of utility as AB units of the public good, or Equivalent Variation (EV) in Figure 1. By asking whether the individual would choose a given increment of the public good or differing amounts of money, the method of paired comparison can pin down WTA within two different dollar amounts (e.g., a non-parametric approach) or estimate WTA using a parametric model such as logit.

Also illustrated in Figure 1 is the more traditional measure, WTP for the gain in AB units of the public good. This involves asking the individual the maximum amount of numeraire they would sacrifice

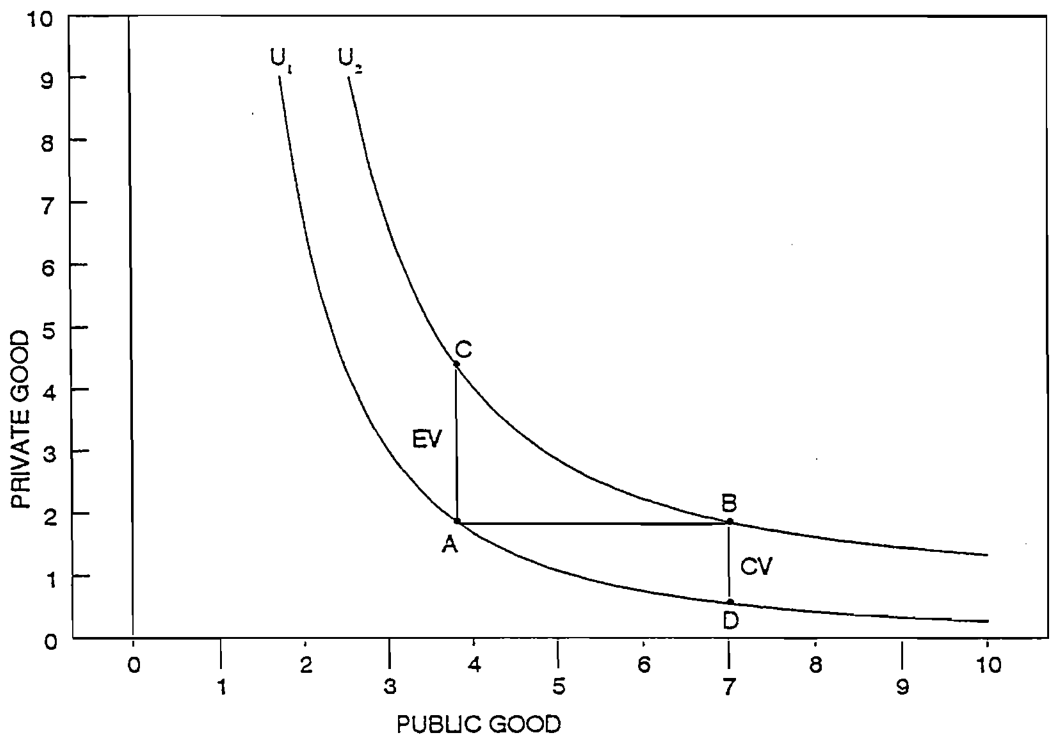


FIGURE 1

to obtain AB, holding utility constant at the original level (U_1). This is Compensating Variation (CV), an amount equal to BD. In this example, BD < CA, therefore CV<EV. While this suggests WTA>WTP, the difference is not expected to be as large as past experiments because the chooser reference point apparently avoids loss aversion and related effects in elicitation of WTA. Using the chooser reference point, any observed difference between WTA and WTP should be related to the income effect.

METHOD OF PAIRED COMPARISON

The psychological theory and method of paired comparison goes back to Fechner (1860) and has developed along several lines. The theory and method of scaling by the "law of comparative judgment" was formalized by Thurston (1927) and further developed and applied by others. David (1988) and Kendall and Gibbons (1990) provide rigorous treatment of the probability theory of comparative judgment. The related methods of categorical judgment (i.e., rating scales, semantic differentials, and rank order methods, etc.) and magnitude and ratio estimation have received similar rigorous theoretical analysis and empirical application in the psychometric literature.

There are at least two conceptual advantages of PC relative to traditional or single bound DC CVM. First, is that an individual is asked to value one good within the context of a bundle of goods. The number of goods in the bundle and the type of competing goods can be varied by the researcher to make the respondent aware of the policy relevant trade-offs. If the government can only provide one or two public goods or services, the choice set could include these possibilities. This should make valuation of the good of interest reflect the presence of the other choices. Most CVM surveys consider just one good, or at most three (Hoehn and Loomis, 1993). Second, the repetitive choices between different dollar amounts and the good provide the opportunity to bracket the individual's valuation between a lower and upper dollar amount. This allows for more precise valuation than is possible with a single bound DC CVM (Hanemann, et al., 1991), although there is some concern about how to statistically analyze multiple responses from the same individual (Cameron and Quiggin, 1994 and Alberini, 1995).

TREATMENTS AND HYPOTHESIS TESTS

The laboratory experiment reported here involves three independent treatments for a private, deliverable good:

#1: Dichotomous choice (DC) WTP in the hypothetical (hyp) payment situation

#2: Dichotomous choice WTP in the real cash payment situation

#3: Paired Comparison (PC) WTA in the hypothetical payment situation

Since our deliverable good is a signed (but not limited edition) commercially available art print, it represents a small fraction of most households' income and is a good with multiple substitutes. Therefore, consumer theory predicts equality of WTA and WTP. The null hypotheses to be tested are:

H^1_o: WTP(Cash-DC) = WTA(Hyp-PC)

H^2_o: WTP(Hyp-DC) = WTA(Hyp-PC)

If we reject H^1_o and H^2_o and find that WTP < WTA, a practical issue remains whether the ratio of WTA (Hyp-PC) to WTP(Hyp-DC) is smaller than has been found in past studies. This would indicate that the method of paired comparison may represent an improvement over standard DC-CVM and merits further research.

These two hypotheses will be tested by comparing whether the 95% confidence intervals on the estimates of mean and median WTP and WTA overlap.

STATISTICAL ANALYSIS

Estimating the Logit Equation and Calculating WTP and WTA

Maximum WTP is not directly observed in the CVM-DC approach nor is minimum WTA directly observed in the PC method. For CVM-DC, there are two basic approaches to estimating maximum WTP: Hanemann's (1984) utility difference approach and Cameron's (1988) compensation function. McConnell (1990) has shown that the linear utility difference model and compensation approach are generally equivalent, and so we adopt Hanemann's as a matter of computational convenience. Hanemann (1984) views CVM respondents using a utility difference approach when they decide whether to answer 'yes' or 'no' at the stated bid amount ($BID). If the utility difference is logistically distributed, a logit model of the probability of a YES response is related to the respondent's bid amount ($BID) and attitude/demographic variables (**Z**) as in equation (1):

(1) $\log[\text{Prob(YES)}/(1-\text{Prob(YES)})] = Bo + B_1(\$BID) + B_2(Z_1) + \ldots + B_n(Z_n)$.

WTP is the area under the cumulative distribution function (CDF or g($BID)) between zero and infinity:

(2) $WTP = \int_0^\infty [1-g(\$BID)]\, d\$BID$ when $WTP > 0$

To calculate the mean WTP from the truncated logistic distribution the formula for the mean of a non-negative random variable is used (Hanemann, 1989:1059):

(3) Mean $WTP = 1/B_1 * (\ln(1+\exp(Bo + (\mathbf{B}_i(\mathbf{Z}_i)))))$

The median is provided by:

(4) Median $WTP = (Bo + (\mathbf{B}_i(\mathbf{Z}_i)))/B_1$

where **Bi** is the vector of coefficients associated with the attitude and demographic variables and **Zi** is a vector of sample means of the associated independent variables.

Calculation of WTA from PC data can be approached in at least two ways. First, there is a non-parametric approach. This approach involves a weighted linear interpolation between the lowest amount the individual said she would accept and highest amount she would not accept. For example, an individual may have said she would not accept $120 for the art print but would accept $180 (and dollar amounts higher than this). Since the interval where she switched from preferring the art print to the money is between $120 and $180, we know WTA lies within this interval. Rather than simply using the mid-point we rely on additional choice information in the paired comparison data to calculate where in this interval the individual's minimum WTA is likely to be. Specifically, we calculate a weighted mid-point using the number of times the individual chooses this good over the other goods in the choice set and over the different dollar amounts (i.e., the rank order of the good over the other goods and monetary amounts). In our example, if \$120 was chosen 6 times, the good was chosen over other goods 7 times and \$180 was chosen 9 times, the weighted interpolation would calculate WTA at \$140 rather than the pure midpoint of the interval (\$150) since the rank order of the good (7) is closer to the rank order of \$120 (6) than it is to \$180 (9). This process of weighted interpolation is repeated for each individual and then a sample mean, median and standard error are calculated. It should be noted that the median calculated from the results of the interpolation approach will likely be different from that estimated from equation (4) when there are small samples such as ours. Further, the median calculated from the logit model is a fitted estimate. The accuracy of such an estimate will depend on how well the logit equation fits the data at the .5 probability level. Lastly, the non-parametric approach also ignores the stochastic nature of the choice process that is reflected in the parametric approach. Specifically, the parametric approach decomposes the explanation of the individual's observed choice into a deterministic component that is observable to the researcher and a stochastic element, that is unobservable to the researcher and treated as random.

The parametric approach to estimating minimum WTA from the PC data allows inclusion of covariates and explicitly incorporates the deterministic and stochastic elements. The approach was developed by Welsh and Bishop (1993) and is called the multiple bounded method. Just as with the interpolation method, each individual's responses are scanned to find the two dollar amounts where the individual switched from a **no** (N) would not accept that amount of money to a **yes** (Y), would accept the money instead of the good. As shown below, there are essentially three possible outcomes: (a)Y_l;

(b) $N_{(\$x)}Y_{(\$z)}$; (c) N_u. Category (a) arises when the individual chooses the lowest amount of money offered (1 or \$4 in our experiment) over the art print; category (b) is where the individual's WTA is bracketed between the highest dollar amount the respondent rejected in favor of the art print (\$x) and the lowest amount they would accept (\$z), where $\$x<\$z$; in category (c) the individual prefers the art print to the highest dollar amount, which was \$295 in this experiment. Assuming the signed art print was not repulsive to the individual, response category (a) is bracketed from below by zero (i.e., if offered the print or zero dollars, they would take the print) and by $4. This bracketing along the real number line is illustrated below:

|------------|--------------------|---------------------------------|--------------------------|----------->

0 $4		$x	$z	$295
	lowest $ amt	highest $ amt rejected	lowest $ amt accepted	highest $ amt
(a)<--Y_l-->		<---(b)$N_{(\$x)}$ $Y_{(\$z)}$ --->		(c) N_u----->

The only difficulty is dealing with response category (c) where the respondent states she would not accept the highest dollar amount offered over the good. This makes it difficult to observe an upper bound on the individual's WTA. However, we do know, with probability =1, that the respondent's WTA is larger than the upper dollar amount. Welsh and Bishop (1993: 339-340) use this observation to program the log likelihood function for this response category.

Using a multiple bounded approach to calculate a sample average WTA involves summing the estimated probability density function over the interval where the individuals response lies.

The log likelihood function is:

$$(5)\ \ln(\text{Likelihood}) = \sum_{i=1}^{n} \ln(Pi_{(\$x)} - Pi_{(\$z)})$$

where, $Pi_{(\$x)}$ and $Pi_{(\$z)}$ are the probabilities that respondent i would reject $x and accept $z, respectively, and n is the number of respondents.

For ease in computing the log likelihood function, the probability density function of WTA is often assumed to be logistically distributed. The log likelihood function is maximized with respect to the vector of parameters (**B**) explaining the pattern of responses observed using a Gauss program developed by Welsh and Bishop (1993). At a minimum the parameters include the bid amount the individual is asked to accept. Additional parameters may include responses to attitude questions or the respondent's demographic characteristics such as age and education. Specifically, the log likelihood function is maximized with respect to **B** as shown in equation (6):

$$(6)\quad \frac{\partial \ln(\text{Likelihood})}{\partial \mathbf{B}} = \sum_{i=1}^{n} \frac{1}{Pi_{(\$x)}-Pi_{(\$z)}} * \left[\frac{\partial Pi(\$x)}{\partial B} - \frac{\partial Pi(\$z)}{\partial B} \right] = 0$$

Using the coefficients estimated in equation (6), mean and median WTA can be calculated from equations (3) and (4), respectively.

Statistically Testing Differences Between WTP and WTA

The equality of WTP and WTA is tested using three independent samples. To compare WTP(cash-DC) and WTA(hyp-PC) we will estimate and compare confidence intervals based on the approach of Park, et al. (1991). If the confidence intervals do not overlap, we conclude that WTA and WTP are different. The advantage of this approach is that it can be applied to the interpolated data as well.

DATA COLLECTION PROCEDURES

Participants

College clerical and administrative staff in academic and non-academic units were recruited and paid $20 for attending one of the 45 minute sessions held on campus. The sessions were conducted before work, at lunch and after work. The Hyp-DC treatment had a total sample of 52 people. The cash-DC treatment had 55 participants. The Hyp-PC experiment involved 103 individuals in 14 sessions (the sessions were smaller due to the limited number of laptop computers available).

Nature of the Comparison Good

While CVM and PC would typically be applied to estimate public or governmentally provided goods, a private good was chosen in this study. First, for the cash treatments the good hadto be deliverable and portable enough that the winning bidder could easily take it with them. To minimize the likelihood that the respondent would simply try to use the market price in determining whether to answer yes or no, a good was needed that was infrequently purchased and sold primarily in specialty stores so that most people would not be familiar with the market price as well as one for which there is a fair amount of price dispersion in the market. Third, we desired a good that had readily observable characteristics, to minimize ambiguity about the product.

Given these desired characteristics, we choose a signed wildlife art print as our good. Art prints can range in price from a few dollars to several hundred dollars, the full extent ofthe product is completely observable, and art often elicits a wide variety of reactions. From among several selections of wildlife art, a signed print of a wolf standing in the forest was selected based on university staff responses to a short questionnaire. The purchase price of the print was $35.

Several pre-test sessions were conducted with university staff to fine-tune the Hyp-DC and Hyp-PC procedures. From these sessions, revisions were made to procedures and instructions until we were satisfied that respondents would understand the task before them in each treatment.

Paired Comparison

Paired comparison can be applied to valuation by specifying a choice set that consists of carefully defined public or private goods (including one or more target goods of special interest) and sums of money. The presentation of a series of paired comparisons between a good and a sum of money was automated by means of an interactive computer program that presents pairs of goods (or a good and a sum of money) from the chooser reference point and requires the respondent to make a choice. To control for order effects, the program presented the binary choices in random order to each individual. The computer program automatically records for each respondent an ordered matrix of binary choices, the sequence in which the respondent sees the pairs and the number of circular triads produced by the respondent's choices.[1] With four goods and ten sums of money and excluding choices between sums of money, the respondent makes 46 choices.

Description and Display of Goods in Paired Comparison Experiments

Four private goods, including the art print, were described on a "product sheet" that was given to each participant. For two of the goods, the wolf art print and phone, the actual goods were displayed in the room. For the other two goods, the restaurant dinner and football tickets, the actual menu and football tickets (along with upcoming football schedule) were mounted on poster boards. The boards and goods were shown up close to the participants and then remained on display during the session.

Structure and Conduct of the Paired Comparison Sessions

The experiments were run with 6-9 people per session. All sessions were led by the principal investigator who followed a written script. The investigator led the participants through the experiment and provided instructions on using the computers. The basic format of each session involved the paired comparison exercise, followed by debriefing questions and finally socioeconomic questions. The entire experiment lasted about 40-50 minutes and was performed on the laptop computers.

Every individual that began the session completed the session, although they were told (in writing) they could leave at any time. Participants were careful to follow directions and did not discusstheir choices with others during the experiment. Observation of participants suggested they put a great deal of thought into their choices. Comments after the session suggested they were stimulated by the experience.

The exact wording of the introduction to the paired comparison choice process was:

---

[1]A circular triad is the product of inconsistent choice. Preference for A over B, B over C, and C over A produces a circular triad (i.e., A>B>C>A). On face value a circular triad implies failure of the transitivity axiom of utility theory. As well understood by psychologists and probability theoreticians, however, the apparent intransitivity may be random or systematic. Random intransitivity, which occurs when the elements in a pair are too similar for consistent discrimination, does not violate the transitivity axiom in expected outcome.

When the choice appears on the screen, please choose the one that you would like to receive if it were to be actually offered to you. Consider each choice independently, as if it were the only choice you had to make. While these choices are hypothetical and you will not actually receive either of the goods, make your choices as if you would actually receive one of the two goods.

Dichotomous Choice WTP

The wording of the Hyp-DC was:

You are being asked to participate in a ***hypothetical*** *sealed bid auction for this art print. We would like to know if you would pay the dollar amount in question #4 below 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 whether you would honestly be prepared to pay the dollar amount stated in question #4 below 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. Please take into consideration your budget and what you can afford to pay. If the price in question #4 is different from what you judge a fair price to be, that is OK. We want to know if you would actually be prepared to pay the price listed in question #4 for the art print.

Take a few moments to think about whether you honestly would be prepared to pay the printed dollar amount for this art print if it were being offered for sale to you today. 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 be required to pay the printed dollar amount. If only one person answers YES, he or she would have obtained the print at the stated price on the survey. If there is more than one person stating YES we will have additional questions to determine who would have been the highest bidder.

4. Would you really be prepared to pay $BID for this art print?

_______ YES, I would pay this amount. _______ NO, I would not pay this amount.

The prelude to the WTP question is different from those of most past CVM questions (particularly those dealing with market goods) in that we asked respondents not to simply estimate what they think the good sells for and to act as if the commitment to pay was real. These two statements were included after debriefing sessions following pretests revealed that respondents were using different criteria to answer the hypothetical as opposed to the real cash WTP questions.

The wording in the Cash-DC question was: *We are now going to conduct a real auction. If you wish to actually buy the art print at the price stated below, answer YES in question #4. If you are the only person who answers YES, you will be required to buy the art print at the stated price. If there is more than one person stating YES, we will have additional questions to determine the highest bidder. 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.*

In any case, the successful buyer will be able to take the art print with them at the end of this session. Now take a few moments to think about what having this art print would be worth to you. If you want to buy the art print at the stated price on the sheet, answer YES. If you don't want to purchase the art print at this price, answer NO.

4. Are you prepared to pay $BID for this art print?

_______ *YES, I will pay this amount.* _______ *NO, I will not pay this amount.*

Dollar Bid Amounts in the Dichotomous Choice and Paired Comparison

In both the Hyp-DC and Cash-DC each person's answer sheet contained one of ten different prices ranging from $2 to $120, but centered around the mean of the pre-test open-ended WTP responses, $38. In the PC, the distribution of bids was similar except the lowest amount was $4 and the highest was $295.

The range of bids for the PC experiment was increased based on the results of the CVM-DC experiments which were conducted prior to the PC experiment.

STATISTICAL RESULTS

Comparison of Demographics Across Sessions

Before establishing whether any differences between treatments were due to differences in stimulus, it is necessary to test whether the respective samples were significantly different or not in terms of standard demographics such as age, education and income. To test for this across our three treatments we performed one-way ANOVA's for education ($F=.74$, $p=.48$), age ($F=2.89$, $p=.06$) and income ($F=.88$, $p=.42$). As indicated by the p values, the samples are not significantly different at the .05 level,although age would be significantly different at the .1 level.

Binary Logit Equations for WTP(Cash-DC) and WTP(Hyp-DC)

We hypothesized that WTP for the art print was positively related to how strongly respondents agreed with the statement that they were in the market for this type of art print (MARKET). This variable had response categories that ranged from 1-5 where 1 is not in the market and 5is strongly agree they were in the market. The MARKET variable had a mean of 3.2 and 2.9 in the Hyp-DC and Cash-DC treatments, respectively. How strongly they liked the art print (LIKE), was also rated on a 5 point scale, with 5 being they strongly agreed they liked the print. Income (INC) measured in thousands of dollars and AGE of the respondent were included as demographic variables. The dependent variable YPAY, is the log of the odds ratio (log[Prob(YES)/(1-Prob(YES))]). Equations (7) and (8) give the logit equations for hypothetical and cash payments:

(7) YPAY(hyp) = -10.77 -.2578($BID) + 1.96(MARKET) +7.84(LIKE)-.537(AGE)+.09(INC)

(t statistics) (-1.75) (2.38) (1.85) (2.27) (-2.12) (1.55)

This logit equation's goodness of fit statistic, the chi-square, equals 56.6 which is significant at the .01 level.

(8) YPAY(cash) = -7.92 - .1787($BID) +1.44(MARKET) +1.37(LIKE)-.04(AGE)+.05(INC)

(t statistics) (2.05) (2.56) (2.47) (1.88) (-.88) (1.36)

This logit equation's chi-square equals 36.6 which is significant at the .01 level.

Logit regressions (7) and (8) indicate that $BID is significant and negatively related to the probability of a "yes" response in both hypothetical and actual markets, whereas being in the market and liking the good increased the probability of a "yes" response.

Multiple Bounded Logit Equations for WTA(Hyp-PC)

The same basic variable specification was used to analyze the paired comparison data using the multiple bounded logit model. The specification included the dollar amount they were asked to accept ($BID), as well as income (INC), AGE and MARKET. Due to the way the multiple bounded logit model is programmed, the dependent variable is coded as zero if the respondent did not choose the print and 1 if the respondent choose to accept the print instead of the dollar amount. Thus, as the dollar amount offered rises, the odds of accepting the print goes down. Individuals whose responses to the series of $BIDS contained circular triads (i.e., they agreed to accept a low amount of money instead of the art print and yet when offered a higher amount of money, choose the art print) were dropped from the paired comparison analysis. These circular triads may arise because the higher amount was offered first and then the lower amount or simply because the point of indifference between money and the art print had been reached causing the individual to randomly switch choices. Nonetheless, the multiple bounded likelihood function cannot handle such inconsistencies as the individual appears simultaneously in several bid intervals, rather than just one. There were 24 individuals out of 103 responses or about 23% of participants that had circular triads for the art print.

Equation (9) provides the coefficients and t-statistics of the multiple bounded logit equation:

(9) ACCEPT PRINT= -2.47	-.0285($BID)	+.393(MARKET)	+.078(AGE)	-.015(INC)	
(t statistics) (-2.63)	(-9.11)	(3.84)	(3.21)	(-1.84)	

The higher the dollar amount offered, the less likely the individual would accept the print instead of the money. The more strongly the individual agreed they were in the MARKET for the art print and the older they were, the more likely they would choose the art print over the money.

Method of Interpolation and Calculation of Standard Errors

After screening to eliminate circular triads, the weighted mid-point of the bid intervals that bracketed an individual's WTA was calculated. Five of the 79 respondents indicated they would choose the art print rather than the highest dollar amount ($295). These five observations were conservatively set at $297. Once each individual's WTA was calculated, the sample mean, median and 95% confidence intervals were computed. These are shown in Table 1.

RESULTS OF HYPOTHESIS TESTS

Table 1 presents the means, medians and 95% confidence intervals for WTP calculated from the dichotomous choice responses and WTA calculated using the two different methods for the paired comparison. It is worth noting that median WTA estimated using the method of interpolation from the PC data ($34) is nearly identical to the (unknown to the respondent) actual purchase price of the signed art print ($35). The difference between the median WTA calculated by interpolation and calculated from the multiple bounded logit model is likely due to the small sample sizes and the fact that the logit model produces an estimated median. For consistency with WTP-DC estimated from the logit model, the comparisons below rely primarily on the estimated WTA from the multiple bounded logit model.

In terms of the first hypothesis, mean WTA(Hyp-PC) exceeds actual cash WTP(Cash-DC) by a factor of 5 using the multiple bounded logit estimate and a factor of 6 using the interpolation estimate (although, the ratio would have been greater than 6 if WTA of the five observations had been truncated at a value greater than $297). The non-overlapping confidence interval suggests these differences are significantly different at conventional levels. Since this ratio encompasses both differences between hypothetical and real commitments as well as WTA vs WTP it is not surprisingthat it is large. Nonetheless, this ratio is smaller than many ratios of either hypothetical/actual or WTA/WTP found in the literature. In particular, the Welsh (1986:153) study is one of the few the elicits both hypothetical and actual WTA *and* WTP using the dichotomous choice question format. He found WTA(CVM-DC)/WTP(Cash-DC) to be 14 in the 1984 Sandhill deer dichotomous choice experiments. It appears that WTA elicited using paired comparison (WTA(Hyp-PC)) produces smaller divergences from WTP(cash-DC) than WTA(CVM-DC) does, but obviously more replications are necessary to determine if this result is robust.

In terms of our second hypothesis, mean WTA(Hyp-PC) calculated using either the parametric multiple bounded and the non-parametric interpolation exceeds WTP(Hyp-DC). Thus, we reject hypothesis number two. The WTA(Hyp-PC) exceeds WTP(Hyp-DC) by a factor of 2.1 ($59/$28) for the mean estimated from the multiple bounded logit model and by a factor of at least 2.4 ($66/$28) for the interpolation approach to estimating WTA. The confidence intervals on WTA(Hyp-PC) using either calculation method do not overlap the WTP(Hyp-DC) confidence intervals. Thus mean WTA(Hyp-PC) is significantly higher than mean WTP(Hyp-DC). The difference in the median WTA(Hyp-PC) and median WTP(Hyp-DC) are smaller and using the interpolation method, the estimates are quite similar (i.e., $34 for WTA(Hyp-PC) vs $28 for WTP(CVM-DC)). Thus, the disparity is reduced when comparing the median.

While the ratio of mean WTA(Hyp-PC) to WTP(Hyp-DC) ranges from 2 to 2.4, this ratio is less than those reported in most other studies. As summarized by Kahneman et al. (1990:1327), hypothetical mean WTA/hypothetical mean WTP is in the range of 2.6-16, averaging 5.38 across the seven studies cited. The performance of our median is even more encouraging with our median WTA/median WTP being 1.2 to 1.75 in our study compared to 3.5 in the three studies cited by Kahneman, et al. 1990.

In fact, the ratio of WTA(Hyp-PC) to WTP(Hyp-DC) is about the same value as the ratio other researchers find using deliverable goods and actual cash. In particular, Boyce, et al., found the ratio of WTA(cash)/WTP(cash) to range from 1.66 to 2.36. Kahneman, et al.'s (1990) summary of four previous experiments had an average ratio of 4.5 for WTA(cash)/WTP(cash).

While there are just a small number of studies for comparison, this pattern of results suggests the possibility that the method of paired comparison may be a promising approach for providing more conservative estimates of hypothetical WTA than CVM-DC.

CONCLUSIONS AND FURTHER RESEARCH

This study has presented an alternative approach to estimate an individual's willingness to accept. A chooser reference point is taken so that the estimate of willingness to accept avoids the loss aversion or endowment effects that apparently elevate estimates of willingness to accept beyond the income effect. The case study application to valuation of a wildlife art print suggests that themethod of paired comparison does provide estimates of hypothetical WTA that are closer in magnitude to hypothetical WTP than found in most of the past WTA/WTP experiments. The ratio of hypothetical WTA estimated using the method of paired comparison to WTP estimated using dichotomous choice CVM was closer to the ratios obtained in experiments where real cash changed hands. Thus, our exploratory case study suggests that the method of paired comparison may represent a promising approach to measuring WTA. We hope these findings stimulate further research in this area by our colleagues.

REFERENCES

Alberini, Anna. 1995. Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models. *Journal of Environmental Economics and Management* 29(2):169-180.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner and Howard Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register* 58(10):4602-4614.

Bishop, Richard and Thomas Heberlein. 1979. Measuring Values of Extramarket Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics* 61(5):926-930.

Boyce, Rebecca, Thomas Brown, Gary McClelland, George Peterson and William Schulze. 1992. An Experimental Examination of Intrinsic Values as a Source of WTA-WTP Disparity. *American Economic Review* 82(5):1366-1373.

Cameron, Trudy. 1988. A New Paradigm for Valuing Non-Market goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression. *Journal of Environmental Economics and Management* 15:355-380.

Cameron, Trudy and John Quiggin. 1994. Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-up" Questionnaire. *Journal of Environmental Economics and Management* 27:218-234.

David, H.A. 1988. *The Method of Paired Comparisons*. Oxford University Press, NY.

Davis, Robert. 1963. The Value of Outdoor Recreation: An Economic Study of the Maine Woods. Ph.D. dissertation, Harvard University.

Fechner, G. T. 1860. Elemente der Psychophysik. Breitkopf and Hartel, Leipzig.

Hammack, Judd and Gardner Brown, Jr. 1974. Waterfowl and Wetlands: Toward Bioeconomic Analysis. Resources for the Future, Washington DC.

Hanemann, Michael. 1985 or6. *Northeastern Journal of Agricultural and Resource Economics*.

Hanemann, Michael. 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics* 66(3): 332-341.

Hanemann, Michael. 1989. Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Reply. *American Journal of Agricultural Economics* 71(4): 1057-1061.

Hanemann, Michael, John Loomis and Barbara Kanninen. 1991. Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, 73(4):1255-1263.

Hanemann, Michael. 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? *American Economic Review* 81(3):635-647.

Hoehn, John and John Loomis. 1993. Substitution Effects in the Contingent Valuation of Multiple Environmental Programs. *Journal of Environmental Economics and Management* 25(1):56-75.

Kahnemann, Daniel, Jack Knetsch and Richard Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98(6):1325-1348.

Kendall, M. and J. Gibbons. 1990. *Rank Correlation Methods*. Fifth Edition. Edward Arnold, Hodder and Stoughton Limited, London.

Lucero, Beatrice. 1996. A Reanalysis of the Empirical Evidence Regarding Willingness to Accept and Willingness to Pay and Its Theoretical Consistency. Ph.D. dissertation, Colorado State University, Fort Collins, CO.

McConnell, Kenneth. 1990. Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation. *Journal of Environmental Economics and Management* 18:19-35.

Mitchell, Robert and Richard Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington DC.

Park, Timothy, John Loomis and Michael Creel. Confidence Intervals for Evaluating Benefit Estimates from Dichotomous Choice Contingent Valuation Studies. *Land Economics* 67(1991): 64-73.

Poe, Gregory, Eric Severance-Loss in, and Michael Welsh. 1994. Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach. *American Journal of Agricultural Economics* 76(4): 904-915.

Thurston, L. 1927. A Law of Comparative Judgement. *Psychological Review* 34:273-286.

Welsh, Michael. 1986. Exploring the Accuracy of the Contingent Valuation Method: Comparisons with Simulated Markets. Ph.D. dissertation, University of Wisconsin.

Welsh, Michael and Richard Bishop. 1993. Multiple Bounded Discrete Choice Models. in Bergstrom, J., Compiler. W-133 Benefits and Costs in Natural Resource Planning, Sixth Interim Report. Dept. of Agricultural and Applied Economics, University of Georgia, Athens, GA.

Table 1.
Hypothetical and Actual WTP from CVM and Hypothetical WTA from Paired Comparison

Treatment	Sample Size	Mean (Median) WTP or WTA by Treatment [95% CI of Mean]			
		WTP		WTA	
		Dichotomous Choice		Paired Comparison	
		Hypothetical Payment	Real Payment	Hypothetical MB Logit	Hypothetical Interpolation
1	52	28 (28) [20-37]			
2	55		11 (9) [6-22]		
3	79			59(52) [39-66]	66(34) [49-83]

MODELING RECREATION DEMAND IN A POISSON SYSTEM OF EQUATIONS: AN ANALYSIS OF THE IMPACT OF INTERNATIONAL EXCHANGE RATES [1] [2]

Jeffrey Englin

Peter Boxall

and

David Watson

ABSTRACT

This paper extends count data travel cost modelling by developing a Poisson system of equations approach to examining recreation demand. During estimation the cross-equation restrictions required to make the system utility theoretic are applied. The model is applied to individual wilderness recreation trips in a system of four Canadian wilderness parks. The resulting demand system is used to examine the impacts of changing American-Canadian currency exchange rates on the participation and welfare of American recreationists.

June 1996

[1] Authors' affiliations are respectively: Department of Applied Economics and Statistics, University of Nevada, Reno; Northern Forest Centre, Canadian Forest Service, Edmonton; Western Field Economics, Edmonton.

[2] Funding for this work was provided by the Northern Ontario Development Agreement, the Canada-Manitoba Partnership Agreement in Forestry, and the Nevada Experiment Station. Scott Shonkwiler and participants in the 1996 W133 Regional Project Meeting provided many helpful suggestions. The authors gratefully acknowledge Janet Baker, Kent Baylis, Dan Benoit and Jennifer Lidgett for research assistance, and John McQueen, Grant Williamson, Rod Drew and Trucia Howard for other support services.

INTRODUCTION

One of the oldest travel cost techniques is the multiple site travel cost method developed by Burt and Brewer. In that analysis Burt and Brewer formalized the idea that several recreation sites could be substitutes for one another in a system of demand equations. This intuition was formalized in two ways. First, the prices of the substitute sites were introduced into each site's demand equation. Second, in order to ensure integrability the demand equations were treated as a system through the imposition of the appropriate econometric cross equation restrictions. Integrability is important whenever the management policies or other considerations such as international exchange rates may change more than a single price.

A recent innovation in travel cost modeling of recreation site demand has been the use of discrete count distributions (Creel and Loomis, Grogger and Carson, Hellerstein). Count distributions are attractive because they allow only nonnegative integer values including zero, which are two common characteristics of recreation demand. These approaches require the use of a semi-logarithmic functional form which requires the application of different integrability restrictions than does a linear demand system (LaFrance).

Previous work by Ozuna and Gomez applied the Poisson distribution to demand systems using the econometric framework developed by King. There are at least two important issues associated with Ozuna and Gomez's application. First, King's econometric framework is based on the bivariate Poisson distribution developed by Holgate that only allows the analysis of two demand equations. Second, the theoretical restrictions provided by Ozuna and Gomez are consistent with *linear* demand system, but are not consistent with the theoretical restrictions required for *semi-logarithmic* demand systems. LaFrance and LaFrance and Hanemann show that the restrictions required to enforce a utility theoretic structure in semi-logarithmic demand systems are substantially different than those used in linear demand systems.

This paper develops a utility theoretic Poisson travel cost demand system and applies it to a census of back country recreation trips taken to a network of four wilderness parks in central Canada. The Poisson system is especially attractive in this empirical example because few recreationists took more than one or two trips to parks in the system in a given year (Boxall et al. 1996a). An additional consideration is that these data are truncated. In our analysis we develop an estimator for truncated *systems* of count equations.

The model is used to simulate the effects of changing currency exchange rates between the United States and Canada on the demand for wilderness trips to four Canadian wilderness parks which are adjacent to each other. The utility theoretic demand system is important in this analysis because the travel costs associated with each park will change simultaneously as exchange rates change. This consideration is significant because the relative travel costs to the parks will change dramatically with exchange rate changes due to the composition of the travel costs to each park. For example, one of the parks is a fly-in backcountry area with the flight costs denominated in Canadian dollars. A system approach to modelling these trips will account for the multiple simultaneous price changes caused by currency exchange rate shifts.

In the next section the theoretical restrictions and econometric procedures used in the analysis are developed. Following that the recreation data used in the empirical application is described and then the results are presented. The paper closes with some conclusions and limitations of this study as well as a discussion of some implications for future research.

METHODOLOGICAL APPROACH

We begin with a demand function for a single site that is specified with a semi-logarithmic functional form. While this functional form is used extensively by applied economists, the recent widespread use of count demand models has made it an attractive functional form for recreation demand estimation (e.g. Shaw, Hellerstein, Grogger and Carson, Englin and Shonkwiler). The particular specification usually used in these studies is:

$$\ln(q_i) = \alpha_1 + \sum_{j=1}^{n} \beta_j P_{ij} + \gamma m_i \,, \tag{1}$$

where q_i is quantity of trips by individual i to site j=1, α_1 is the intercept associated with site j=1, P_{ij} are travel costs faced by i for trips to j=1 and all other sites (j=2,...,n), m_i is individual i's income and β_j and γ are parameters to be estimated. Consumer surplus associated with a season's recreation trips for the ith individual is found by integrating (1) between two prices P_0 and P_1. If only a single site is of interest, (1) is sufficient for welfare estimation. If, however, the concern is valuing recreation at a system of sites, then a system of demand functions must be estimated and integrability becomes an important consideration.

Theoretical Considerations

LaFrance and LaFrance and Hanemann demonstrate the conditions that a system of semi-logarithmic demand functions must fulfill to form an integrable demand system. These conditions take the form of restrictions on the intercept, cross-price effects and the income effect in the system of equations. To illustrate these consider the following system of demand functions for trips to 4 sites:

$$\begin{aligned}
\ln(q_{i,j=1}) &= \alpha_1 + \sum_{j=1}^{4} \beta_j p_{ij} + \gamma_1 m_i \\
\ln(q_{i,j=2}) &= \alpha_2 + \sum_{j=1}^{4} \beta_j p_{ij} + \gamma_2 m_i \\
\ln(q_{i,j=3}) &= \alpha_3 + \sum_{j=1}^{4} \beta_j p_{ij} + \gamma_3 m_i \\
\ln(q_{i,j=4}) &= \alpha_4 + \sum_{j=1}^{4} \beta_j p_{ij} + \gamma_4 m_i \,.
\end{aligned} \tag{2}$$

The n-1 intercept restrictions are:

$$\alpha_j = \alpha_1 \left(\frac{\beta_{jj}}{\beta_{11}}\right), \tag{3}$$

where α_j is the intercept for the jth site, β_{jj} is the own price coefficient for the jth site, and α_1 and β_{11} are the intercept and own price coefficient for site 1. The effect of this restriction is that only one of the intercepts, α_1, is estimated in the econometric model. The remaining intercepts are calculated as functions of α_1 and the two own price parameters as shown in (3). The second restriction is that there is only one income effect (γ) for the system. In essence, the sub-utility function that describes the closely related sites has a single income effect rather than one for each site.

Finally, the uncompensated cross price effects are all restricted to be zero. Note, however, that the compensated cross price effects are non-zero (see Shonkwiler). These are calculated by using the Slutsky equation to identify them as follows:

$$s_{ijk} = q_{ij} \frac{\partial q_{ik}}{\partial m_i} = \gamma q_{ik} q_{ij} \tag{4}$$

where s_{ijk} is the compensated substitution effect between sites j and k for individual i, and the q's are the quantities of trips to the sites in the system by individual i. Notice that the cross price effects will be symmetric (i.e. $s_{jk} = s_{kj}$) for individual i, but will not be identical across individuals who may choose different quantity pairs. Thus, the compensated semi-logarithmic system parameters that are calculated for any individual may appear like a cross-price constrained incomplete linear demand system. However, this is misleading. In the linear system the *parameters are constant* regardless of the individual's consumption point. In the semi-logarithmic system the *relationships among the parameters* are constant across individuals.

As a result of these restrictions only a single intercept (α_1), a single income effect (γ), the own price parameters (β_{jj}), and any parameters associated with variables used when individuals are pooled need to be estimated.[1] A particular point at which to evaluate the economic relationships described above needs to be chosen to recover the full compensated demand system implied by these parameters. Possible candidates include the average or the median individual in the sample. An alternative approach taken in this study is to calculate the implied compensated demand system for each individual in the data and take the mean of the results.

These theoretical conditions are sharply different from the development of Ozuna and Gomez. Ozuna and Gomez do not restrict the intercept, include the uncompensated cross-price effects, or constrain those cross-price effects to be symmetric. This presentation and the work of LaFrance and LaFrance and Hanemann show that the Ozuna and Gomez specification is not utility theoretic in a semi-logarithmic demand system framework. The restrictions imposed in their analysis *are* consistent with the linear demand framework (see Burt and Brewer).

Econometric Considerations

The empirical application of the model described below, involves a Poisson demand system which is consistent with the semi-log functional form. The estimation is accomplished by specifying a log likelihood function that includes a Poisson equation for the demands at each site. The likelihood for n sites can be written as:

$$\prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{q_i}}{q_i!}, \tag{5}$$

where the latent quantity demanded, λ, is $\exp(X\theta)$, the X's represent prices, income and other variables, θ's represent parameters (γ, β) to be estimated, and the cross equation restrictions presented in the theoretical section are applied to each set of parameters.

A common difficulty in applied work is that trip data are truncated. Trips are only observed when the number of trips to at least one site in the system is positive. The visitation to the other sites in the system may be zero. In these situations a truncated version of (5) must be developed. Following Grogger and Carson the general relationship between the untruncated and truncated models can be written as:

$$f_0(q_i) = \frac{f(q_i)}{1 - F(0)}, \qquad (6)$$

where $f_0(q_i)$ is the truncated density, $f(q_i)$ is the probability function, and F(0) is the distribution evaluated at 0. In the case of a system of n Poisson demand functions can be written as:

$$\prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{q_i}}{q_i!} \left(1 - \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{0}}{0!}\right)^{-1}, \qquad (7)$$

or, upon some evaluation,

$$\frac{\prod_{i=1}^{n} e^{-\lambda_i} \lambda^{q_i}}{q_i!} \left(1 - \prod_{i=1}^{n} e^{-\lambda_i}\right)^{-1}. \qquad (8)$$

After further simplification the likelihood function can be written:

$$\prod_{i=1}^{n} \frac{\lambda_i^{q_i}}{q_i!} \left(\prod_{i=1}^{n} e^{\lambda_i} - 1\right)^{-1} . \tag{9}$$

Taking logarithms of (9) results in the log likelihood function used to estimate the truncated system.

DATA

The recreation demand data used to illustrate the model come from Nopiming, Whiteshell and Atikaki Provincial Parks in Manitoba and Woodland Caribou Provincial Park in Ontario. The parks are adjacent to each other and close to the Canada-USA border (Fig. 1). Unlike the Boundary Waters Canoe Area (BWCA) in Minnesota or Quetico Provincial Park (QPP) in Ontario, these parks maintain no entry restrictions or quotas. Backcountry use fees are absent in the three Manitoba parks, while a small daily use fee is charged at the Ontario park and a small entry fee at Whiteshell park. Recreation in these parks is not highly regulated and currently involves small numbers of people in comparison to the more popular BWCA and QPP.

Trip data used in developing the model come from self-registration of visitors. The data we analyze includes Canadian and US residents who visited the four parks during 1993 and 1994 (see Watson et al. and Boxall et al. 1996a, 1996b). The registration information includes the name and address of the group leader, group member's names, group size, trip length in days and boat type, among other things. This information was entered into a computer database and registrants' names and addresses were cross-linked to determine multiple visits in the two years to one or more of the four parks. This resulted in a sample of 1122 trips by 939 groups to the four park system during the two years.

For the 939 individuals in the database, 812 took one trip, 95 two trips, 21 three trips and 11 four or more trips to the parks during the two year period (Boxall et al. 1996a). Most of the visits were to Nopiming park, while the fewest was to Atikaki. Multiple trips were found to include more than one trip to the same park as well as trips to multiple parks. The Poisson count framework should be especially attractive in analyzing these trips because few individuals took more than one or two trips to a park and many visited only a single park.

Travel costs were based on four components: i) the out-of-pocket expenses for vehicle travel, estimated at $0.22 CDN/km; ii) the value of travel time which was based on an average speed of 80km/hr and the opportunity cost valued at 1/3 the wage rate which was estimated using average incomes from each census zone and an assumption that individuals worked 2080 hours per year; and iii) other costs such as payment for commercial float plane access to backcountry areas (only in Atikaki and Woodland Caribou Parks). For each person in a recreation group, we assumed that the out-of-pocket travel costs and float plane costs were split among members of the group, but that each individual had to pay their own travel time costs. The only applicable fees in these parks are a $5.00 entry fee per group for Whiteshell Park and a per person daily use fee of $5.00 for Woodland Caribou Park. Float plane access usually cost a flat rate of $477.80 plus $25.00 per person from Bisset, Manitoba, where float plane services originated (Boxall et al 1996a).

Distances from the group leaders' home town to each park were measured by the shortest highway route with a planimeter and 1:250,000 scale maps of Manitoba and northwestern Ontario and Canadian highway maps. United States distances were measured using ZIPFIP from the home location to the border crossing consistent with minimizing the total distance travelled. Each park was assigned a common single entry point in this estimation. For US recreation groups we distinguished the distances travelled in Canada and the US so that only the Canadian portion would be subject to currency exchange differences.

Other information obtained from the permits and survey forms was used to estimate income, m_i, or other variables. The availability of an individual's postal code or zip code allowed the estimation of socio-economic data using the most recent national censuses. For Canadian visitors this was obtained from the 1991 Canada Census (Statistics Canada, 1993, CD-ROM version) and for US visitors the information was obtained from ZIPFIP (Hellerstein et al.). The socio-economic data included average household income, average education level and average household size. All estimates provided in US dollars were converted to Canadian dollars based on the Bank of Canada official exchange rates for 1994 ($1.366 CDN = $1.00 USA). Thus, variables representing an individual's household income, education and household size. A dummy variable for having a US residence was constructed (i.e. the variable=1 if the person lived in the US and equalled 0 if in Canada).

RESULTS

Table 1 presents maximum likelihood estimates of parameters of the untruncated and truncated models. For both models all travel cost coefficients are negative and significant well beyond the 1% level. While all socioeconomic variables were tried in the analysis, only the USA dummy variables are significant. The positive signs on two of them suggest that Atikaki and Woodland Caribou Parks were preferred over Nopiming and Whiteshell Parks for wilderness trips in the system by USA recreationists. This finding corresponds to the visitation levels by groups from the USA in the database (Boxall et al. 1996a). The income effect is positive and insignificant. The truncated model has a lower log likelihood value than the untruncated model.

Table 2 summarizes trip information and provides basic welfare results for each model. The first three rows show the average travel cost, group size and trip lengths by wilderness park. These data suggest that Nopiming and Whiteshell parks are dominated by weekend trips and that Atikaki and Woodland Caribou provide, on average, week-long trips. Per trip consumer surplus estimates vary with the treatment of truncation - the untruncated model generates consumer surplus estimates that are about 2-3 times larger than those from the truncated model. The welfare estimates for Whiteshell and

Nopiming Provincial Parks are much smaller than those for trips to Atikaki and Woodland Caribou Parks.

The scale of the welfare estimate associated with wilderness recreation at the four parks can be seen by dividing the per trip welfare by the average days per trip. The consumer surplus per day ranges from $107 per day at Nopiming for the truncated model to $1754 per day for Atikaki for the untruncated estimate. The untruncated values seem particularly high. Since the truncated model has a lower log likelihood and matches the data better than the untruncated model we focus on the truncated welfare results. However, two factors should be considered in the comparison of welfare estimates. First, there is virtually no economic research on Canadian wilderness recreation that provides nonmarket values for comparison. Wilderness trips in the United States are considerably different than Canada due to the scale of use, the presence of more stringent management regulations, and higher levels of congestion (e.g. Boxall et al. 1996a). Thus, comparisons with US recreational values may be problematic.

Second, Atikaki and Woodland Caribou parks, which are remote and require considerable effort and skill to access unless float plane services are hired, provide much greater recreational values than Nopiming and Whiteshell which primarily provide weekend trips for local Canadian residents. Trips to Atikaki and Woodland Caribou parks are probably more similar to high quality ecotourism trips than standard wilderness experiences. Thus, the benefit estimates reported here should be compared with those types of recreation estimates in the literature.

To fully illustrate the results the individual truncated model is expanded to show an entire compensated demand system using the income coefficient. The parameters of this full system are shown in Table 3. The parameters reported were found by calculating the full compensated system for each individual in the survey and then averaging those estimates. Ten of the 28 resulting coefficients reported in Table 3 are repeated from Table 1. These include the own price parameters, the intercept on Nopiming Provincial Park, the income parameter and the USA shift parameters. The cross price effects and the intercept terms for Atikaki, Whiteshell and Woodland Caribou Provincial Parks were calculated using (3) and (4) which were presented in the second section of this paper.

The compensated demand system presented in Table 3 looks like a linear demand system with cross price symmetry imposed on the estimates. In this case, however, the specific cross price parameters are dependent on the quantities at which the system was evaluated. Since the cross price terms are all calculated as $0.0041*q_k*q_j$ it is possible that some cross price effects will be zero which will occur if no individual took trips to both sites k and j. In these data the cross price effects between Whiteshell Park and the two most remote parks, Atikaki and Woodland Caribou are zero. Since Whiteshell park maintains the highest levels of development in the four park system, this finding is not unexpected and suggests that subsets of the parks cater to different recreation markets.

We use these econometric results to examine changes in the exchange rate between Canada and the United States. Table 4 shows the results for a variety of exchange rates. The first column shows the number of $US needed to buy a $1 CDN. These range from $0.50 to par ($1.00). The next four columns show the effect of changes in exchange rates on the visitation by American recreationists. The model predicts that Nopiming will have the greatest reduction in American visitors. The effects range from 3.95% for a modest five cent increase in the value of the Canadian dollar, to a 23.37% drop if the two currencies become equal. The decline in American visits for the other parks is more modest. The change in exchange rates has the least impact on Whiteshell. This is because there are small fees (denominated in Canadian dollars) to enter the park and the park is the closest of the four to the United States. The two remote parks also see little impact on visitation. This is primarily because the demand for these parks is inelastic. There are no substitutes for these parks in the continental United States.

The final column of Table 4 reports the change in American welfare associated with changes in exchange rates.[2] The welfare effects are total seasonal effects that account for welfare impacts across the entire system. The welfare effects range from $134.95 per season for a five cent change in exchange rates to nearly $800 per season if the currencies are at par. A comparison with Table 2 shows that the losses are a small portion of the per trip values suggested by the model.

SUMMARY

This paper develops Poisson demand system models both with and without truncation. The model was applied to wilderness recreation trips to 4 Canadian wilderness parks. Using the truncated model the value of a day of wilderness recreation in these parks appears to be substantial. Trips to the most remote of the parks was estimated to be worth over $700 per day.

An important policy finding of this analysis is that a stronger Canadian currency will have an strong effect on visitation to only one of the parks, Nopiming Provincial Park. The other three parks are either very close to the United States (Whiteshell) or have poor substitutes in the United States (Atikaki and Woodland Caribou). As a result, the American demand for these parks is inelastic, and therefore largely unresponsive to a stronger Canadian currency.

ENDNOTES

1. The appendix shows LaFrance and Hanemann's indirect utility function, expenditure function and the formula for equivalent variation associated with three alternative specifications of the income effect. These correspond to positive, negative and zero income effects. The selection between these models is an empirical question that must be decided on a case by case basis.

2. The welfare calculations are made using the zero income effects equivalent variation formula provided in the appendix. If we use the equivalent variation formula associated with the positive income effect four members of the sample have undefined welfare measures. This is due to the small average *census* income associated with their postal code. Using the zero income effects version (since the income effect is insignificant) provides well defined welfare measures for all members of the sample.

BIBLIOGRAPHY

Boxall, P., D. Watson and J. Englin. Backcountry Recreation in Central Canadian Wilderness Parks: A Systems Approach. Northern Ontario Development Agreement Technical Report, Canadian Forest Service, Sault Ste Marie, Ontario. 1996a.

Boxall, P., D.O. Watson and J. Englin. "Backcountry Recreationists' Valuation of Forest and Park Management Features in the Canadian Shield Region." *Canadian Journal of Forest Research*. 26(1996b):in press.

Burt, O. and D. Brewer. "Estimation of Net Social Benefits from Outdoor Recreation." *Econometrica*. 52(1971):32-41.

Creel, M.D. and J.B. Loomis. "Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California." *American Journal of Agricultural Economics*. 72(1990):434-441.

Englin, J. and J. S. Shonkwiler. "Estimating Social Welfare in Count Models." *Review of Economics and Statistics*. 77(1995):104-112.

Grogger, J.T. and R.T. Carson. "Models for Truncated Counts." *Journal of Applied Econometrics*. 6(1991):225-238.

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., D. Woo, D. McCollum and D. Donnelly. 1993. ZIPFIP: A ZIP and FIPS Database Users Manual Economic Research Service, United States Department of Agriculture in cooperation with United States Forest Service, Washington D.C.

Holgate, P. "Estimation for the bivariate Poisson distribution." *Biometrika*. 51(1964):241-245.

King, G. "A Seemingly Unrelated Poisson Regression Model." *Sociological Methods and Research*. 17(1989):235-255.

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

LaFrance, J. and M. Hanemann. "On the Integration of Some Common Demand Systems." Staff Papers in Economics. Agricultural Economics and Economics Department, Montana State University, Bozeman, MT. 1984.

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.

Shaw, Daigee. "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation and Endogenous Stratification." *Journal of Econometrics*. 37(1988):211-223.

Shonkwiler, J.S. "Systems of Travel Cost Models of Recreation Demand." Paper presented at the W133 Regional Project Meetings. Monterey, California. March 1995.

Watson,D., L. Peters, P. Boxall, J. Englin and K. Chakraborty. The economic value of canoeing in Nopiming Park in relation to forest and park management. A report of the 1993 field season: Volume I. Canadian Forest Service Report, Canada-Manitoba Partnership Agreement in Forestry, Winnipeg Manitoba. 1994.

Table 1. Maximum likelihood estimates of parameters for untruncated and truncated recreation demand system models for four Canadian wilderness parks.[a]

	Model Specification	
	Untruncated	**Truncated**
Intercept	-0.2774*** (-2.02)	-1.5803*** (-4.88)
Nopiming travel cost	-0.0013*** (-2.94)	-0.0034*** (-5.14)
Atikaki travel cost	-0.00008*** (-2.76)	-0.0002*** (-4.25)
Whiteshell travel cost	-0.0005*** (-3.02)	-0.0012*** (-5.20)
Woodland travel cost	-0.0001*** (-2.84)	-0.0002*** (-4.57)
Nopiming USA Dummy	-1.4011*** (-5.50)	-0.3924 (-1.31)
Atikaki USA Dummy	1.7935*** (8.07)	2.5165*** (9.35)
Whiteshell USA Dummy	-1.7513*** (-4.22)	-0.9685*** (-2.20)
Woodland USA Dummy	2.6703*** (15.77)	3.3836*** (14.84)
Income	0.0016 (0.55)	0.0041 (0.57)
Log Likelihood	-2108.68	-1528.07

[a] - t statistics are provided in the parentheses.

*** - significant at the 1% level or beyond.

Table 2. Average use, group size, travel costs, and consumer surplus associated with two alternative specifications of demand for a system of four Canadian wilderness parks.

	Wilderness Park			
	Nopiming	**Atikaki**	**Whiteshell**	**Woodland Caribou**
Mean days/trip	2.72	6.45	2.50	6.00
Mean group size/trip	4.03	5.93	4.41	3.81
Mean travel cost ($) per person	102.57	389.30	102.14	203.15
Consumer surplus/trip				
Untruncated	758.54	11312.6	1969.46	9451.47
Truncated	293.23	4941.66	871.65	4176.85
Consumer surplus/day				
Untruncated	278.87	1753.90	787.78	1578.25
Truncated	107.80	766.14	327.06	696.14

Table 3. Implied compensated demand parameters for the truncated individual demand system.

Variable	**Nopiming Park**	**Atikaki Park**	**Whiteshell Park**	**Woodland Caribou Park**
Intercept	-1.5803	-4.4047	-2.6057	-4.2366
Price Coefficients				
Nopiming Park	-0.0034	0.0745	0.0657	0.0438
Atikaki Park	0.0745	-0.0002	0	0.0175
Whiteshell Park	0.0657	0	-0.0012	0
Woodland Caribou	0.0438	0.0175	0	-0.0002
Demand Shifters				
Income	0.0041	0.0041	0.0041	0.0041
USA dummy	-0.3924	2.5165	-0.9685	3.3836

Table 4. Percentage change in trips by park and overall welfare impacts on US recreationists by simulating changes in US-Canadian exchange rates.

	Percent change in trips				
Exchange Rate $US per $CDN	**Nopiming Park**	**Atikaki Park**	**Whiteshell Park**	**Woodland Caribou Park**	**Average US Welfare Effects**
0.50	15.67%	1.21%	0.07%	0.60%	$547.22
0.55	11.79%	0.91%	0.05%	0.45%	$409.28
0.60	7.88%	0.60%	0.03%	0.30%	$272.11
0.65	3.95%	0.30%	0.01%	0.01%	$135.68
0.70[1]	0.00%	0.00%	0.00%	0.00%	$0.00
0.75	-3.95%	-0.30%	-0.01%	-0.15%	-$134.95
0.80	-7.90%	-0.61%	-0.03%	-0.30%	-$269.17
0.85	-11.82%	-0.91%	-0.05%	-0.40%	-$402.68
0.90	-15.72%	-1.22%	-0.07%	-0.60%	-$535.47
0.95	-19.57%	-1.52%	-0.09%	-0.70%	-$667.57
1.00	-23.37%	-1.83%	-0.11%	-0.90%	-$798.97

[1] this rate approximates the exchange rate during the time the data were collected (1993 and 1994).

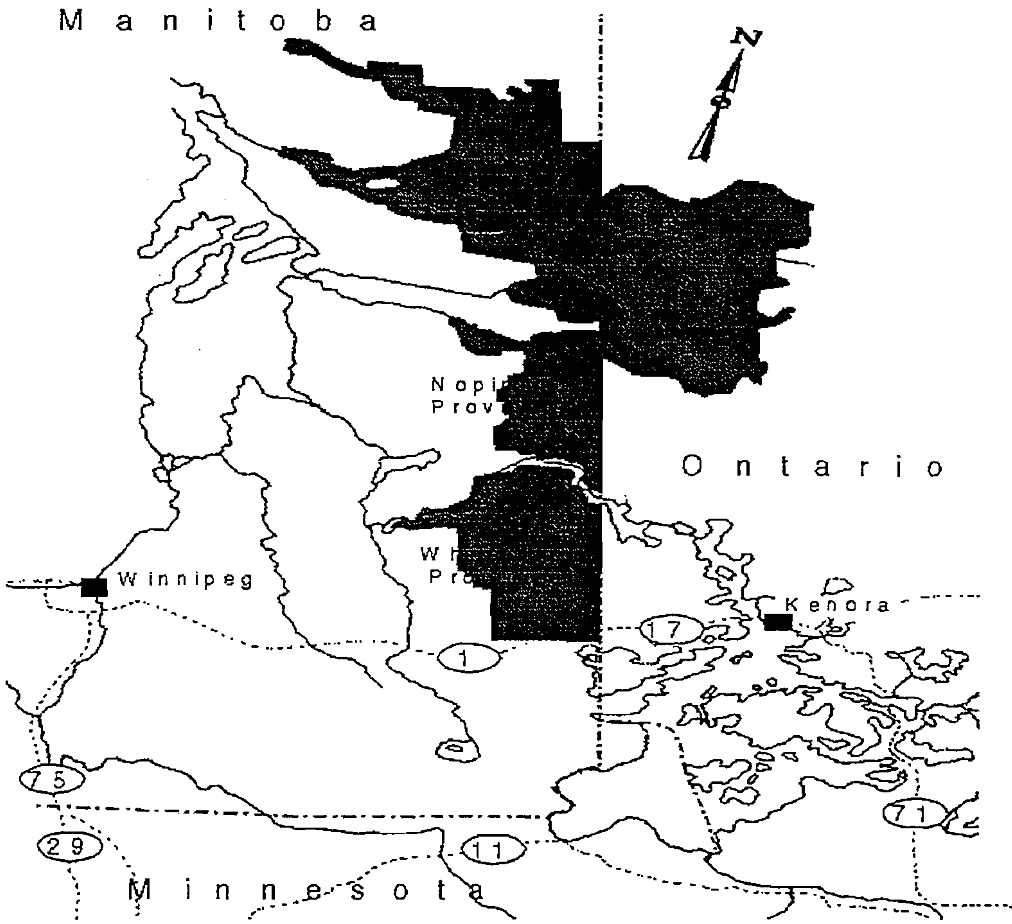


Figure 1. Map of the Four Canadian Wilderness Parks

APPENDIX

The quasi-indirect utility functions, quasi-expenditure functions and formulae for the equivalent variation for three different income effects for the semi-logarithmic demand system used above taken from LaFrance and Hanemann.

	Quasi - Indirect Utility Function	Quasi - Expenditure Function	Equivalent Variation
Positive Income Effects	$(-\frac{1}{\gamma})e^{-\gamma m} - (\alpha_1/\beta_{11})e^{\sum_{i=1}^{k}\beta_{ii}p_i} - \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i}$	$(-\frac{1}{\gamma})\ell n[-\gamma(\theta + (\frac{\alpha_1}{\beta_{11}})e^{\sum_{i=1}^{n}\beta_i p_i} + \sum_{i=k+1}^{\ell}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i} + \sum_{i=\ell+1}^{n}\alpha_i p_i)]$	$(-\frac{1}{\gamma})\ell n[e^{-\gamma m} - \gamma((\frac{\alpha_1}{\beta_{11}})(e^{\sum_{i=1}^{k}\beta_{ii}p_i^0} - e^{\sum_{i=1}^{k}\beta_{ii}p_i^1}) + \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})(e^{\beta_{ii}p_i^0} - e^{\beta_{ii}p_i^1}))] - m$
Negative Income Effects	$(-\frac{1}{\gamma})e^{-\gamma m} - (\frac{\alpha_1}{\beta_{11}})e^{\sum_{i=1}^{k}\beta_{ii}p_i} - \sum_{i=k+1}^{\ell}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i} - \sum_{i=\ell+1}^{n}\alpha_i p_i$	$(-\frac{1}{\gamma})\ell n[-\gamma(\theta + (\frac{\alpha_1}{\beta_{11}})e^{\sum_{i=1}^{n}\beta_i p_i} + \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i})]$	$(-\frac{1}{\gamma})\ell n[e^{-\gamma m} - \gamma((\frac{\alpha_1}{\beta_{11}})(e^{\sum_{i=1}^{k}\beta_{ii}p_i^0} - e^{\sum_{i=1}^{k}\beta_{ii}p_i^1}) + \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})(e^{\beta_{ii}p_i^0} - e^{\beta_{ii}p_i^1}) + \sum_{i=\ell+1}^{n}(p_i^0 - p_i^1))] -$
Zero Income Effects	$m - (\frac{\alpha_1}{\beta_{11}})e^{\sum_{i=1}^{k}\beta_{ii}p_i} - \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i}$	$(\frac{\alpha_1}{\beta_{11}})e^{\sum_{i=1}^{n}\beta_i p_i} + \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})e^{\beta_{ii}p_i} + \theta$	$(\frac{\alpha_1}{\beta_{11}})(e^{\sum_{i=1}^{k}\beta_{ii}p_i^0} - e^{\sum_{i=1}^{k}\beta_{ii}p_i^1}) + \sum_{i=k+1}^{n}(\frac{\alpha_i}{\beta_{ii}})(e^{\beta_{ii}p_i^0} - e^{\beta_{ii}p_i^1})$

QUANTILE METHODS OF USING COUNT DATA MODELS IN TRAVEL DEMAND ESTIMATION

Daniel Hellerstein
USDA/ERS/NRED/EIRA

5/15/96

ABSTRACT

The use of count data models in travel cost analysis has become widely accepted. Although it has been established that count models provide consistent estimates of expected demand, the prediction of demand for known individuals has received little attention. In this paper, we will look at a quantile method of using count models to make individual specific predictions. This method allows accurate prediction of demand, and consumer surplus, for individuals for whom actual demand data exists. Such knowledge may be useful when accurate cost benefit measures are desired for projects affecting specific, readily measured, groups.

INTRODUCTION

Early advocates of the use of count models (Mullahy, Cameron and Trivedi, Smith, Shaw) focused mainly on their appealing econometric properties, such as consistent control for censoring, and the recognition of the integer only status of the dependent variable. If only implicitly, the assumption was that the functional forms being estimated were the result of some textbook-standard constrained optimization process, with a random variable (ϵ) included to account for otherwise unexplainable stochastic results.

While convenient, this assumption is weak. Clearly, the non-integer nature of travel demand is fundamental, and not just a statistical complication. Recognizing this, later workers (Hellerstein and Mendelsohn) developed individual demand generating processes that would yield quantities that follows a count distribution. This count distribution could then be used to make statements about the expectation of quantity demanded (E[Y]), as well as statements about the expectation of consumer surplus generated by the good (E[CS]). Fortunately, it can be shown that the E[Y] and E[CS] generated by these behaviorally based models is identical to the "naive" functions (borrowed from the non-count literature) commonly used in earlier applications of count models.

While useful, these results only apply to expectations. If one desires to make statements about a set of observed individuals, say should their environment change from E_0 to E_1, expectations may provide only a rough gauge. If information on an individual's observed demand (under E_0) can be used, more accurate predictions of actual demand (under E_1) should be obtainable (that is, more accurate than the expectation under E_1).

Consider the analogous case of continuous, linear demand with additive error term. Ignoring issues of censoring, OLS will provide an unbiased estimate of the quantity demanded under current conditions ($\hat{y}|E_0$). The residual ($y-\hat{y}=\hat{e}$) between quantity demanded and observed demand is also an unbiased estimate of the value of an individuals random variable (ϵ). To predict demand under E_1, one simply computes the expectation of demand ($\hat{y} | E_1$), and adds $\hat{e}$.

For count models the problem is more difficult. It is not at all clear how one would extract $\hat{y}$ and $\hat{e}$ from a count model. It has been suggested (Haab and McConnell) that given some expectation (say, $E[Y]=\lambda=\exp(X\beta)$ in a Poisson model), observed demand will be the result of a multiplicative process involving λ and a highly constrained (over it's range of support) error process. While mathematically feasible, such stories require very unusual error structures, and offer little guidance as to what may occur should λ change (that is, how does the range of support of ϵ change as λ changes).

Instead, this paper recommends adopting what shall be called a "quantile" based approach. The key assumption is that for an individual i, given a properly specified count demand (say, the Poisson with parameter λ_i) and an observed demand (say, Y_i), the analyst can state what "quantile of relative demand" the individual occupies. Assuming that this quantile does not change as the shape of the distribution changes (say, as λ_i decreases), a measurable change in the distribution will yield a readily computed, unique value of Y_i. Furthermore, by integrating under estimates of Y_i, an individual specific consumer surplus can be computed.

MODEL

Without loss of generality, we assume that an individual i's trip demand is Poisson distributed with parameter λ_i:

1) $$f(n|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^n}{n!} \quad ; \quad n=0,1,..$$

where: λ_i = $\exp(X_i\beta)$
X_i = Observable variables for individual i
β = Known coefficients.

Note that equation (1) summarizes our knowledge of what might occur, given values of X_i and β, but is mute concerning the mechanism that generates the observed randomly distributed results.[1] Instead, given X_i (implying a known λ_i) and a known quantity demand (Y_i), an individual "quantile of relative-avidity" (κ_i) can be determined, with κ_i a two element vector defined as:

2)
$$\begin{aligned}
&\kappa_i[1] \equiv \kappa[1 \mid Y_i, \lambda_i] = F(Y_i - 1 \mid \lambda_i) \equiv \textit{(lower bound on avidity)} \\
&\kappa_i[2] \equiv \kappa[2 \mid Y_i, \lambda_i] = F(Y_i \mid \lambda_i) \equiv \textit{(upper bound on avidity)} \\
&\textit{where} \\
&\quad F(J \mid \lambda) \textit{ is the CDF, defined as: } \sum_{j=0}^{J} f(j \mid \lambda), \\
&\quad \textit{and } \kappa[1 \mid 0, \lambda] \equiv 0.0, \ \forall \lambda.
\end{aligned}$$

κ implies that a $\kappa_i[1]$ fraction of a population of similar individuals[2] will have demand less then Y_i, and a $1\text{-}\kappa_i[2]$ fraction will have demand greater then Y_i.[3] That is, a $\kappa_i[1]$ fraction of this population are less avid users of the resource, and a $1\text{-}\kappa_i[2]$ fraction are more avid. Hence, the actual (but unobservable) level of "relative" avidity (K_i) is a value between these endpoints.

With equation 2 in mind, the following postulate forms the crux of this paper:

Postulate 1: *Should λ_i change, κ_i will not change.*

[1]For example, one might postulate the existence of an unobservable multiplicate factor, or one might postulate the operation of a sequence of discrete choices. The model offers no guidance as to which is true, so long as each generates identical results.

[2]That is, a population of individuals with identical X and β, hence identical values of λ.

[3] Of course, a fraction $f(Y_i \mid \lambda_i)$ will demand Y_i.

In words, postulate 1 states that the κ of an individual is not influenced by changes in the level of demand; that changes in λ do not cause a reordering of the "avidity" ranking.[4]

Postulate 1 is analogous to the condition of zero correlation between explanatory variables and the additive random error term (Johnston, pp 172), required for unbiased linear regression. In both cases, the assumption is that knowledge of observable features gives no information on the relative magnitude of the error influences.[5] In fact, one could view the random error term of linear regressions as a shorthand for a point κ, with $\kappa[1] \cong \kappa[2] \cong K$.

In the context of the Poisson model, the use of κ for predictive purposes is predicated on postulate 1, in combination with a proposed new value of λ:

Postulate 2: *Given:*

a) a known λ_a *and* Y_a,

b) $\kappa_a = \kappa_a[i=1,2] = \kappa[i \mid Y_a, \lambda_a]$

c) a new λ_b,

d) a candidate prediction of demand Y_b, *and*

e) $\kappa_b[i=1,2] = \kappa[i \mid Y_b, \lambda_b]$;

then Y_b *is permissible only if*

i) $\kappa_b[1] \le \kappa_a[1] \le \kappa_b[2]$

or

ii) $\kappa_a[1] \le \kappa_b[1] \le \kappa_a[2]$.

[4]More precisely, K does not change, and neither does κ; where κ summarizes the information concerning the true location of K.

[5]Note that knowledge of explanatory variables may give some information on the possible "size" of a random factor, but not on it's "sign" (viz. the use of weighted least squares for control of heteroscedasticity).

In words, postulate 2 states that the κ_b implied by a new λ (λ_b) and a proposed quantity (Y_b) must overlap some portion of κ_a. That is, for a candidate Y_b to be observed, there must be a (not necessarily unique) K that lies within both κ_a (i.e.; is consistent with observing Y_a , given λ_a) and lies within κ_b (i.e.; is consistent with observing Y_b, given λ_b).

For continuous models, where κ collapses to K, an exact measure of Y_b can be obtained. With count models, which possess a chunky distribution, in many cases the set of permissible values of Y_b will contain more then one element. In such cases, an expected Y_b should be computed.

Postulate 3: Given λ_a and Y_a, and a new λ_b, the expected value of Y_b is calculated as:

$$E[Y_b] = \sum_{j=0}^{\infty} j * W(j \mid \lambda_b, \kappa_a)$$

where,

$$\kappa_a[i=1,2] = \kappa[i \mid Y_a, \lambda_a)$$

$$\kappa_b[i=1,2] = \kappa[i \mid j, \lambda_b)$$

and W defined using

a) *if* $\kappa_b[2] < \kappa_a[1]$ *then* $W(j \mid \lambda_b, \kappa_a) = 0$

b) *if* $\kappa_b[1] > \kappa_a[2]$ *then* $W(j \mid \lambda_b, \kappa_a) = 0$

c) *if* $\kappa_b[1] < \kappa_a[1]$ *and* $\kappa_a[1] < \kappa_b[2] < \kappa_a[2]$

$$\textit{then } W(j \mid \lambda_b, \kappa_a] = \frac{\kappa_b[2] - \kappa_a[1]}{f(Y_a \mid \lambda_a)}$$

d) *if* $\kappa_a[1] < \kappa_b[1] < \kappa_a[2]$ *and* $\kappa_a[1] < \kappa_b[2] < \kappa_a[2]$

$$\textit{then } W(j \mid \lambda_b, \kappa_a] = \frac{f(j \mid \lambda_b)}{f(Y_a \mid \lambda_a)}$$

e) *if* $\kappa_a[1] < \kappa_b[1] < \kappa_a[2]$ *and* $\kappa_b[2] > \kappa_a[2]$

$$\textit{then } W(j \mid \lambda_b, \kappa_a] = \frac{\kappa_a[2] - \kappa_b[1]}{f(Y_a \mid \lambda_a)}$$

f) *if* $\kappa_b[1] < \kappa_a[1]$ *and* $\kappa_b[2] > \kappa_a[2]$

$$\textit{then } W(j \mid \lambda_b, \kappa_a] = 1.0$$

Note that $\sum W_j = 1.0$

Postulate 3 states that the expected value of Y_b will be a weighted sum of all integers; with non-zero weights for all values j whose quantile ($\kappa_j=\kappa(j,\lambda_b)$) overlap the original quantile ($\kappa_a=\kappa(Y_a,\lambda_a)$). More precisely, this weight equals the fraction of κ_a that lies within κ_j.

Postulate 3 can be readily programmed, yielding predictions of Y_b under new values of λ (λ_b), say as generated by an increase in an X variable (i.e.; an increase in price). This prediction is noted as $Y^*[\lambda_b | \kappa(\lambda_a, Y_a)]$. For illustrative purposes, the appendix contains a short numerical example of these computations.

Furthermore, a measure of individual specific consumer surplus (CS^*) is readily obtained by numerically integrating under Y^* between λ_a and a sufficiently small λ (λ_0). The numeric integration proceeds by slowly increasing price by dp, computing Y^* at the resulting lambda, and summing the resulting dp*Y^* rectangles. Note that λ_0 is the largest solution to $Y^*(\lambda_0 | \kappa(\lambda_a, Y_a))=0$; since Y^* will equal 0 for any smaller value of λ^0, one need not carry the integration beyond this point.

A NUMERICAL EXAMPLE

Results later .. Preliminary analysis indicate that when sample cs is desired (and not a cs after changes in exogenous variables) "naive" method may be more robust than this method.

THIS IS A WORKING PAPER FOR PRESENTATION AT A CONFERENCE DEDICATED TO SUCH WORKS!

REFERENCES

Cameron, Colin and Pravin Trivedi. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests". Journal of Applied Econometrics 1 (Jan. 1986),29-53

Haab, Timothy and Kenneth McConnell. "Count Data Models and the Problem of Zeros in Recreation Demand Analysis", unpublished manuscript, Department of Agricultural Economics, University of Maryland.

Hellerstein, Daniel and Robert Mendelsohn. "A Theoretical Foundation for Count Data Models". American Journal of Agricultural Economics 74 (August 1993), 604-611.

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

Shaw, Dai Gee. "On-Site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification". Journal of Econometrics 37 (Feb. 1988),211-223.

Smith, V. Kerry, "Selection and Recreation Demand". American Journal of Agricultural Economics 70(1988):29-36

Appendix: Numeric Examples of Computing Y^*

In this section we provide a few examples to illustrate the steps needed to compute Y^*.

This table of pdfs (f) and cdfs (F) of the Poisson is helpful in the following examples.

j	f(j,2.6)	F(j,2.6)	f(j,2.1)	F(j,2.1)	f(j,0.6)	F(j,0.6)
0	0.0743	0.0743	0.122	0.122	0.548	0.548
1	0.193	0.267	0.257	0.379	0.329	0.878
2	0.251	0.518	0.270	0.649	0.098	0.976
3	0.217	0.736	0.189	0.838	0.0198	0.996
4	0.141	0.877	0.099	0.937	0.003	0.997

1) $Y^*(2.1|2.6,4)$: the expected value of "observed" Y (Y_0) when λ is 2.1, given that Y_0 equaled 4.0 when λ was 2.6.

From the above, we determine that:

$\kappa[1|2.6,4]=\kappa_a[1]=0.736$, $\kappa[2|2.6,4]=\kappa_a[2]=.877$;
$\kappa[1|2.1,3]=0.649$, $\kappa[2|2.1,3]=.838$; and
$\kappa[1|2.1,4]=0.838$, $\kappa[2|2.1,4]=.937$.
$f(4 | 2.6) = 0.14$

Since $\kappa_a[1]$ falls within the range of $\kappa[2.1,3]$ and $\kappa_a[2]$ falls within the range of $\kappa[2.1,4]$, a fraction (of those with $Y_0=4$ when $\lambda=2.6$) will demand 3, and a fraction will demand 4.

Precisely:

By postulate 3, case c:

$$w_3=(\kappa[2|2.1,3]-\kappa_a[1])/f(4,2.6) = (0.83-0.74)/0.14 = 0.72;$$

by postulate 3, case e:

$$w_4=(\kappa_a[2]-\kappa[1|2.1,4])/f(4,2.6)= (0.877-0.83)/0.14 = 0.27.$$

With w_3 the fraction demanding a quantity of 3, and w_4 the fraction demanding 4, we obtain:

$$Y^*(2.1|2.6,4) = w_3*3 + w_4*4 = 0.72*3 + 0.27*4 = 3.28.$$

(note that rounding errors cause the sum of probabilities to be less the 1.0)

2) $Y^*(0.6|2.6,4)$: Note that $\kappa[2|0.6,1]=0.878$ and $\kappa[1|0.6,1]=.548$. Since $\kappa_a[1]=0.736$ is greater then $\kappa[1|0.6,1]=.548$, and $\kappa_a[2]=0.877$ is less then $\kappa[2|0.6,1]=0.878$. Thus, by case f of postulate 3, $W_1=1$, and we obtain:

$$Y^*(0.6|2.6,4)=1.0 * 1.0 = 1.0.$$

3) All values of λ less then 0.1 will yield $Y^*=0$, since $\kappa[2|0.1,0]=0.904$ is greater then $\kappa_a[2]=0.877$.

TRAVEL COST MODELS OF THE DEMAND FOR ROCK CLIMBING

by

W. Douglass Shaw
Department of Applied Economics and Statistics/204
University of Nevada
Reno, NV 89557-0105
E-mail: wdshaw@unr.edu

and

Paul Jakus
Department of Agricultural Economics and Rural Sociology
P.O. Box 1071
University of Tennessee
Knoxville, TN 37901
E-mail: pjakus@utk.edu

ABSTRACT

In this paper we estimate the demand for rock climbing and calculate welfare measures for changing access to a number of climbs at a climbing area. In addition to the novel recreation application, we extend the travel cost methodology by combining the double hurdle count data model (DH) with a multinomial logit model of site-choice. The combined model allows us to simultaneously explain the decision to participate and allocate trips among sites. The application is to climbers who visit one of the premiere rock climbing areas in the northeastern United States and its important substitute sites. We also estimate a conventional welfare measure, which is the maximum WTP rather than lose access to the climbing site.

* Shaw is corresponding author, senior authorship not assigned. The authors thank Glenn Hoagland, Thom Scheuer and other staff members at the Mohonk Preserve for allowing us to add questions to their survey, and Gong Chen, Becky Stephens, and Jae Espey for excellent research assistance. The authors also wish to acknowledge Scott Shonkwiler for his considerable assistance in the empirical research and Trudy Cameron and Jeff Englin for valuable comments on the manuscript. Any errors that remain are our own. Research supported by the Nevada and Tennessee Agricultural Experiment Stations and the W-133 Regional Research Project.

1. INTRODUCTION

Mountain and rock climbing had an estimated 4.2 million participants in the U.S. in 1991, and it is estimated that 100,000 new climbers try some version of the sport each year (The Economist). The rapid growth of climbing has led to proposed rules by the U.S. National Park Service and the Department of Interior that may affect climbing on federal lands. As stated in the Federal Register, "the increased impacts to park resources because of this activity suggest that regulations and guidelines need to be developed to protect park resources..." [Fed. Reg. 58, June 14, 1993]. Despite its growing popularity and the apparent need for new management strategies, there are no published estimates of the basic value of climbing, the impacts of site quality changes, or the proposed regulations on rock climbing. Previous research efforts have focused on *why* individuals become attracted to climbing or on the risk aspects of the sport. Barring the unpublished work by Ekstrand (1994) however, no research has been expressly devoted to <u>economic</u> modeling of the demand for rock climbing or mountaineering.

This paper serves to fill that void. After a description of rock climbing and our data in Section 2, Section 3 presents the three models used to estimate demand for rock climbing–a site-choice model, a trip frequency model and a combined site choice-frequency model. The final model represents an extension of current travel cost methods by combining the site choice model with a double hurdle count data model. We present all three models because of the need to explore differences in welfare estimates from each approach and because there has been little previous work to suggest the most appropriate type of empirical model. In Section 4 we present the empirical demand models and consumer surplus estimates; finally, we summarize the paper and offer suggestions for future research in Section 5.

2. BACKGROUND ON ROCK CLIMBING AND THE DATA

2.1 The Sport of Rock Climbing

Rock climbing differs from "mountain" climbing in that the former most frequently involves climbing a rock cliff in good weather and does not involve negotiating ice and snow. Rock climbers are often interested in a shorter, extremely technical section of the cliff, and their goal of climbing this section in good form is quite different from the mountaineer's goal of reaching a summit. The sport is sometimes construed by the general public as a hazardous activity, but climbers can exercise some control over the risks they personally assume by using the proper equipment and judgement (Jakus and Shaw).

Technical rock climbing on smaller cliffs or "crags" involves the choice of specific routes up the rockface, where routes differ in their degree of difficulty, length, and hazard. Falling is a part of the sport for most climbers, but equipment is used to protect the climber from hitting the ground or the side of the cliff after falling. This equipment varies from metal devices placed permanently in the rock (such as a bolt or piton), to devices which can be temporarily inserted into cracks and fissures, and removed as the climbers advance upward (called chockstones or nuts).[1] As the "leader" climbs using only the features of the rock, the rope is threaded through these devices. Because the second climber holding (belaying) the rope from below, the devices act as potential pivot points in the event of a fall. The climbing equipment is used only to protect against the consequences of a leader's fall which would otherwise result in injury. After belaying the leader, the second advances upward, but he or she is well protected by the rope above.

Climbing routes are subjectively rated according to technical (gymnastic) difficulty and risk. Ratings are published in readily available guidebooks (for popular areas) or spread by word of mouth (for less popular areas). Guidebooks note the location and length of a route, its technical difficulty[2], and whether the climb can be well protected or not (the hazard scale). Many guidebooks feature "maps" of the specific route, showing rock features and permanent protection points.

[1]Recently there has been a growing distinction between climbing areas which primarily offer permanent bolted protection and those which primarily offer temporary protection, requiring the climber to place nuts and chockstones. Areas which offer mostly temporary protection are called "traditional" areas, while areas with permanent protection are called "sport climbing" areas.

[2]The difficulty scale in the U.S. runs from the easiest technical climb at 5.0 to the most difficult, at 5.14. The technical rating is akin to the difficulty rating assigned to dives in diving competitions. Ratings reported in a guidebook are a combination of ratings by experts and feedback from other climbers.

2.2 *The Data*

Relative to other recreationists such as hunters or anglers, it is very difficult to collect data on climbers. An intercept survey method raises objections about whether those intercepted at the site are representative of the general population of climbers (Shaw 1988). A sample drawn from the general population would be extremely costly because most households contain no members that climb. Alternatively, one can find known groups of climbers such as organization members some other way.

Our data were collected using a survey of members of the Mohonk Preserve (MP) in New York state. The Preserve is New York State's largest non-profit nature preserve and is about 65 miles from the New York city metropolitan area. The MP receives a large number of visitors, particularly on good weather weekend days. Visitors can become annual members of the MP (a non-profit organization) by paying an annual fee entitling them to free entry for the year, or they may forego membership and purchase a daily entry pass. Not every Preserve visitor is a climber (many hike, view nature, bike, and do other outdoor activities), but the MP is an international climbing destination and is arguably the most important climbing area in the northeastern United States.[3] Among national climbing areas, it is somewhat unusual in that it offers virtually no bolted climbing.

The MP staff initiated the survey to elicit management preferences from approximately 2,500 members. The survey questionnaire was mailed only once in an envelope along with the Preserve's Fall 1993 newsletter. The survey budget did not allow follow up methods as suggested by Dillman and others. Because of controversial management policies relating to congestion, access, and conflicts between different users, direct WTP questions were excluded from the questionnaire. Eight hundred ninety two usable surveys were obtained.

Of members returning the survey, 220 said they used MP primarily to climb. Trip data were collected from this group of climbers. Information included the number of trips taken to the Preserve in 1993, as well

[3] The climbing cliffs at the Preserve are also known as the "Gunks" and have been featured in an article on climbing in the international publication, The Economist.

as the total number of trips taken to important substitute climbing areas. Usable trip and travel cost data were obtained from 183 respondents. We do not have complete information on each specific trip that each of these 183 climbers took in 1993, and several self-described climbers did not take a climbing trip to any destination in 1993. (In our final estimating sample, almost ten percent of the climbers take zero climbing trips in 1993).

In modelling demand for climbing, we recognize the potential bias in using just a sample of members.[4] There is no way to know how our sample differs from the general climbing population because no data has ever been collected for the latter group. We can, however, compare our mail survey respondent characteristics to those of a separate on-site sample conducted in Fall 1993, which unfortunately does not contain information on the individual's residence location. The mail (members only) sample climbers have similar incomes, age and climbing expenditures to the on-site (non-member) climbers. Members and non-members also visited other northeastern climbing areas in similar patterns. On average, members visited MP more often than non-members (17 trips vs. 5 trips), and there is a higher proportion of males among the members, but this may be due to a higher probability of males to respond to a mail survey. Although we do not infer that our sample is representative of <u>all</u> rock climbers in U.S., we believe the sample could be representative of climbers in the northeast.

2.3 *Measuring Site Characteristics*

Site characteristics are important in modeling the demand for recreational areas, but the travel cost literature does little to aid us in selecting an appropriate site characteristic for rock climbing areas. Instead, we draw on our own experience.[5] We hypothesize that an appropriate characteristic is the number of routes available to the to the climber, where the limiting factor is the individual's technical ability. Technical ability dictates the hardest route level that can be climbed; climbing any harder than one's ability may result in frustrating failure or bodily harm. While climbers *do* sometimes attempt routes harder than current technical

[4] Additional bias introduced by failure to return the questionnaire is also possible (Cameron et al.).

[5] The authors are both climbers, each with over fifteen years of experience.

ability as a means to improve, they most often choose those near their current limit. Climbers also do not generally seek out routes well beneath their ability because these are present little physical or mental challenge. Thus, if a climber can lead 5.10 routes and there are 200 such routes at area *A*, then that is the site characteristic of interest when choosing among sites. Our site characteristic is similar to the ability-specific characteristic Morey constructs for skiers and ski area choice (Morey 1985) and, like Morey, we assume ability is exogenously determined, being based on long run acquired skill through experience, practice, and a climber's natural physical gifts.

3. THE MODELS AND CONSUMER'S SURPLUS

Our data permit several variants of the travel cost model to be estimated, particularly the random utility (RUM) and count data models (see Bockstael, McConnell and Strand (1991) for a recent review of recreation demand models). The RUM and count data models each have limitations. For example, the conventional multinomial logit (MNL) model cannot easily be used to estimate the total number of trips an individual takes in a season and therefore leads to difficulties in estimating a seasonal or annual welfare measure.[6] In contrast, the count-data approach handles seasonal demand for a single recreation site, but cannot be easily used to examine decisions to allocate among two or more sites simultaneously (Shonkwiler 1995). In addition, the single site count data model is not as rich as the RUM in how it incorporates site substitution because of difficulties in correctly specifying the model with cross price terms in a way that is consistent with utility maximization, which again has implications for welfare measures.

Many recent efforts theoretically or econometrically link the total number of trips an individual takes in the recreation season to the choice of a recreation destination on any given trip (Yen and Adamowicz; Hausman, Leonard and McFadden; Terza and Wilson; Parsons and Kealy). Such models rely on mixing a RUM with a trip frequency model to neatly obtain seasonal, rather than per-trip, welfare measures. These

[6]By making several strong assumptions, versions of the "repeated" logit or nested logit models do allow exploration of the participation decision, and allow derivation of seasonal welfare measures (see for example, Morey et al. 1991).

models also allow the individual to adjust total trips taken during the season in response to site quality changes, rather than assuming the individual's total trips stay constant, with possible reallocations among various destinations. The site choice model demands are conditional on the total trips taken, but the latter can be jointly estimated with the former.

Because our application involves a rapidly growing recreation activity demanding new management strategies, we have chosen to employ three models which highlight different dimensions of the demand for climbing and are suitable to meet different policy objectives.

3.1 The Multinomial Logit (MNL) Site Choice Model

The data reveal how often climbers went to the four most important sites throughout the northeastern U.S., so a site-choice model can be estimated. In addition to the Preserve, the three other climbing areas are Ragged Mountain (RM) in Connecticut, the Adirondacks (A) in upstate New York, and the White Mountainsaround Conway, New Hampshire.[7] RM differs from the other three in that it offers only short climbs, virtually all of which may climbed by first taking a trail to the top and then hanging a rope down the cliff.

If the usual assumptions about the distribution of the error vector are made, an MNL model can be estimated via the log likelihood function:

$$\ln \mathcal{L} = \sum_{j=1} y_j \ln \pi_j \tag{1}$$

where the probability of visiting site j is π_j, or:

$$\pi_j = \frac{\exp(X_j\beta)}{\sum_{k=1} \exp(X_k\beta)} \tag{2}$$

[7]Another important climbing area in the northeast is located near Bar Harbor, Maine, so distant from any major population center (except perhaps Boston) that it is not as important as a major destination site.

y_j is the number of trips to site j and X_j is a vector of explanatory variables that explain site-allocation, which can also vary for the individual (i). Surplus measures from an MNL can be calculated as shown in Hanemann (1984). However, the simple multinomial logit model does not allow calculation of seasonal compensating or equivalent variation measures without imposing strong behavioral assumptions.

3.2 Modeling Annual Climbing Trips: Count Data Approaches

The trip-taking data are also well suited for one or more variations of the count model. Count data travel cost models are increasingly popular (Hellerstein; Creel and Loomis; Englin and Shonkwiler). The frequency of the climber's total trips (y) to the MP can be modeled using the basic Poisson model distribution with location parameter λ. λ can be parameterized as:

$$\lambda = \exp(w'\tau) \tag{3}$$

where the vector of variables in w explain the frequency of total trips taken to the MP and τ is the corresponding vector of parameters.

There are many variations on the basic Poisson model. For our purposes, the most important deal with excess zeros (Greene) and, related to the problem of excess zeros, the participation decision (i.e. the decision to enter the market at all). Because our sample of members includes many who do not take a climbing trip to the Preserve, we use a hurdle model, which helps explain the participation decision.

A Double Hurdle Count Data Model

A hurdle mechanism (Mullahy) can be introduced to explain the decision to enter the market (in our case, whether to climb during 1993). The discrete choice double hurdle (DH) Poisson model (as laid out by Shonkwiler and Shaw) allows for two kinds of zero trip takers: those who did not climb anywhere, and those who did climb elsewhere but for some reason did not climb at a specific site like the Preserve. The DH model is consistent with the zero modified Poisson (ZMP) discussed in Johnson and Kotz and is essentially the same as the "zero altered Poisson" (ZAP) discussed by Greene. The model is not the same as the *single* hurdle model of Mullahy's, nor do Johnson and Kotz or Greene explain their models as "double" hurdles (Shonkwiler and Shaw).

Define D_i to be equal to the latent decision to consume trips (desired trips are equal to y^*) so that $y = 0$ if D_i is less than or equal 0. Let the vector of variables that explain participation (go or not) be z, which includes variables describing personal or demographic characteristics (these may or may not include variables in the vector w, which explain trip frequency). Then,

$$E(D_i) = \theta_i = \exp(z_i'\gamma) \tag{4}$$

If trips are positive, then observed consumption equals desired consumption, or:

$$y = y^*, \quad E(y^*)=\lambda \tag{5}$$

where λ is defined in equation (3).

With two hurdles, the outcome of no consumption (non-participation) can be observed for two reasons: the desired consumption is non-positive or, if it is positive, an additional hurdle (D less than or equal to zero) still can prevent participation. If the two hurdles are independent of one another, the Poisson likelihood function for the double hurdle (suppressing the individual subscript i) is:

$$\mathfrak{L} = \prod_{y=0} [\exp(-\lambda)+(1-\exp(-\lambda))\exp(-\theta)]\cdot\prod_{y>0} (1-\exp(-\theta))\exp(-\lambda)\frac{\lambda^y}{y!} \tag{6}$$

The log likelihood for (6) will be assured of being well behaved because the parameterization of θ assures us that $exp(-\theta)$ will lie between zero and one.

The CS measure from the DH count data model reveals the approximate WTP for a trip to a site, rather than lose access to it (Shonkwiler and Shaw). However, if a site characteristic of interest does not significantly explain the hurdle portion of the model, then the value of a characteristic change cannot be isolated. Because the site characteristic likely affects the frequency of visits to the site more than the decision to go at all (the participation hurdle), the DH welfare measure is not likely to be relevant in estimating welfare measures for changes in characteristics.

A Joint Multinomial Logit - Double Hurdle Poisson Model

RUMs rarely are used to model the demand for trips across all sites for an entire season, as RUMs assume that trips to a site are conditional on seasonal trips having been allocated outside of the model. Following the expanding empirical literature (including Terza and Wilson, Yen and Adamowicz (YA), and Hausman, Leonard and McFadden (HLM), we combine the site choice and season's trips models by jointly estimating the multinomial logit and count data models. No prior studies combine the double hurdle (ZMP) with the MNL as we do here. We first develop probabilities of visiting site j, conditioned on participation. Assuming the multinomial distribution for the probabilities of visiting site j conditional on total seasonal trips (t), we have:

$$P(y_1,y_2,\ldots,y_J|t) = t!\prod_{j=1}^{J} \pi_j^{y_j} \;/\; \prod_{j=1}^{J} y_j! \tag{7}$$

where $t = \Sigma y_j$. If we also assume that the π_j stem from a random utility model where the error term follows the extreme value distribution, these conditional probabilities can be specified and estimated using the multinomial logit model, as above. Combining equation (7) with the double hurdle poisson (6) leads to the following joint frequency outcome, denoted MNL-DH (adopting the notation from equations above):

$$g(y_1,y_2,\ldots,y_J) = \exp(-\lambda)+(1-\exp(-\lambda))\exp(-\theta) \quad \textit{for } t = 0 \tag{8}$$

and for positive seasonal trips, $t > 0$,

$$= \frac{(1-\exp(-\theta))\exp(-\lambda)\lambda^{t}\prod_{j=1} \pi_j^{y_j}}{\prod_{j=1} y_j!} \tag{9}$$

Define $d = 1$ for those who take no trips during the season ($t = 0$) and $d = 0$ for those who do ($t > 0$). The joint frequency distribution in (8) and (9) leads to the log likelihood function:

$$\mathcal{L} = (d)\sum_{0} [\ln\{\exp(-w'\tau) + (1-\exp(-w'\tau)\ \exp(-\exp(z'\delta))\}] \tag{10}$$

$$+ (1-d) \sum_{+} [\ln(1-\exp(-z'\delta))-w'\tau + w'\tau \sum_{j}^{J} y_j + \sum_{j}^{J} y_j \ln(\pi_j) - \sum_{j}^{J} \ln(y_j!)]$$

Obtaining the total consumer's surplus in the joint MNL-count data model using the total trip demand equation is consistent with two-stage budgeting (HLM). The CS is the integral under this total trip demand function.

4. EMPIRICAL APPLICATION AND RESULTS

4.1 Specification and Parameter Estimates

Site Choice Model

For the simple MNL model we assume that the explanatory variables include the site's implicit price (travel costs) and the site characteristic. A site-specific intercept term for Ragged Mountain is also included because, as noted previously, this site is different than the other three. Table 1 provides the results of this simple model. Overall results are reasonable – a modified R^2 for the model is approximately .53 – and each variable is significantly different from zero. Because utility is linear in the explanatory variables the sign can be easily interpreted and we note that the number of climbs at the maximum level of the climber has a positive influence on site choice.

Double Hurdle Model

For the double hurdle model we partition the variables into those which explain the frequency and the participation decisions. We assume that the frequency of climbing trips is a function of the site price and the site characteristic. Table 2 provides basic results of the Poisson count with double hurdle model. A modified R^2 shows that the model explains about 31 percent of the variation in total trips. As can be seen in the frequency portion of the table, the price term is negative and significantly different from zero while the characteristic is positive and significant. The constant term captures some systematic positive effect.

The survey was not designed to specifically address the decision to take at least one climbing trip, so there were few variables from which to choose for the participation hurdle.[8] The variables included in the participation hurdle portion are limited to leading ability and a taste variable indicating the importance of the Preserve's environmental education programs they influence the decisions to become a member (an integer from 1 = most important through 5 = least important). Our hypothesis is that, all else equal, climbers who can lead harder climbs are more likely to go climbing at least once than those who focus on environmental education. The variable has the expected influence in the empirical model.

Joint Model

Results from the joint multinomial-Poisson model are presented in Table 3, estimated using full-information maximum likelihood (FIML). The site choice model is specified identically to the MNL model above. The double hurdle specification is also similar to the simple single site double hurdle model above, with one key difference. For the joint model, we must develop a price index for all trips to the four sites under consideration. Following Bockstael et al. (1984) and more recently HLM and others, the price index is the inclusive value from the MNL. The sign of the index parameter, unlike that of a conventional price term, is expected to be positive in the combined model. This is because the index is a preference weighted measure of costs and site characteristics (we note that Parsons and Kealy derive a different index theoretically, splitting the site travel cost and characteristics effects).[9] In our model it is important in the

[8]Shonkwiler and Shaw suggest several possible variables on which to solicit information in a survey questionnaire which may help explain the decision to stay home for the season.

[9]In HLM, the inclusive value term is positive in three of their four specifications, but they obtain the "wrong sign" in one. Readers may be confused because they switch the signs in their table of results. More discussion of the index can be found in Bockstael, McConnell, and Strand (1984), Yen and

frequency, rather than the participation portion, of the double hurdle model. Because most previous authors do not have but one portion of the count data model, this differentiation does not occur.

In the joint model the site choice results are quite robust and parameters have the expected signs. The double hurdle portion of the model, as in the single-site model, is more problematic, but the price index has the expected positive sign. Greater technical ability leads to more annual trips, which is a nice intuitive result. The specification for the participation hurdle was somewhat problematic, as the survey was not designed to elicit variables to explain this, and we were able only to specify this portion with the constant and environmental importance variables. The latter has the expected sign, indicating that the less important the role of environmental education in becoming a member, the more likely the person will take a climbing trip.

4.2 *Estimates of Consumer Surplus*

The focus in this section is on welfare estimates for climbing at the Preserve. While conventional CS measures for access to the MP can be estimated, the more policy-relevant questions are associated with changing the number of available climbs at the Preserve. For example, climbing routes at the Sky Top area are off limits to Preserve climbers during at least part of the season. These climbs are not actually on Preserve grounds and are the property of the Mohonk Hotel, so access to these climbs may become at risk. The seasonal cliff closure at Sky Top is similar to seasonal closures at many other U.S. climbing areas during times when birds of prey nest on cliffsides. Another reason to be interested in the number of available climbs stems from proposed regulations on climbing. The number of climbs at a given area can be increased by permanently bolting new routes. In the United States, federal guidelines banning the use of bolts in National Parks and recreation areas under the jurisdiction of the Department of Interior are interpreted to exist already [CFR 36, § 1,2], and new guidelines have been proposed. Movements to legislate stricter regulations could result in noticeable effects on climbing on federal lands.

Adamowicz, and Terza and Wilson.

It is impossible to do an exact simulation of route access restrictions at Sky Top but, because these routes are included in the site characteristic measure used in the demand models, the loss can be approximated using a percentage decrease in the number of routes available at the MP. The MNL model yields a per-trip estimate of welfare losses while the joint MNL-DHP model yields a seasonal estimate. Estimates for two reductions and two increases (10 and 50 percent) in the site characteristic appear in Table 4. None of the numbers are large, as even the seasonal measure from the joint model yields an individual maximum of \$16.00 loss for a huge (the 50 percent) decrease in the Preserve's climbs, with a sample average of only \$7.85. In both route reduction scenarios, however, climbers still have available over 300 routes and all routes at the three substitute sites. Thus, our evidence suggests that cliff closures and bolt bans do not result in large welfare losses for members when a large set of substitutes is available.

Finally, it is useful to compare the per trip benefits for rock climbing to other benefit estimates for sports such as recreational fishing, hiking, skiing, etc. Our comparable "per-trip" estimates are from the single-site Poisson/double hurdle model. Using the double hurdle model, we obtain benefit estimates in the range of \$70 to \$90 per trip, with the average CV being about \$80. While we recognize that discussion of these should be treated carefully (Morey 1994), the estimate of WTP per-trip is in the range of "per-choice occasion" or "per-trip or per-day" estimates of WTP for special recreation such as fishing for salmon in Alaska.

Only one other recreational rock-climbing study provides results to which ours can be compared. Though his is a mail survey, Ekstrand originally intercepted climbers for his sample at Eldorado Canyon State Park, an internationally known climbing area near Boulder, Colorado. Using four different versions of the travel cost model[10], CS was between \$39.51 and \$48.73 per trip. These estimates were made assuming CS reflects the average climber who is taking only one-day trips.[11] Ekstrand also estimated CS using the

[10]He estimates OLS, truncated OLS, the Poisson, and the negative binomial models.

[11]We caution against too much reliance on Ekstrand's TCM estimates however, because the travel cost functions include total days climbed in a season as right-hand side explanatory variables, but

contingent valuation questions posing current and future simulated conditions. For his current conditions, the CVM approach yields between roughly $11 and $26 per day, depending on whether the WTP obtained from the CVM is adjusted for the opportunity cost of travel time. Because his survey was conducted in 1991, we assume that the CS estimates are in 1991 dollars. Our single site DH CS estimates (in 1993 dollars) are higher than Ekstrand's using any of his methods. As our sample is of members of the Preserve and Ekstrand's is an on-site sample with no adjustment for on-site sample bias, neither may be representative of some climbing population at large.

5. SUMMARY AND CONCLUSIONS

This paper provides the only estimates of a model of the demand for rock climbers other than the unpublished study by Ekstrand (1994). The travel cost methodology has been extended to allow for a double hurdle participation mechanism and for allocation of trips among many sites. We have provided the first estimates of consumers surplus associated with seasonal cliff closures at climbing areas. Except for our conventional measure of annual WTP, welfare effects of various policy scenarios are small. Because of the nature of our sample (many substitute sites, but all offering only traditional climbing) it should <u>not</u> be inferred the general population of climbers is willing to pay only a small amount to prevent loss of existing climbs or bring about bolting of new climbs. The magnitude of welfare losses is probably a function of available substitutes, so more regional studies should be conducted. Until these areas are studied, however, this study contains the only available estimates.

REFERENCES

Anonymous author. 1995. Climbing up the wall. (The) Economist (March 11).

Bockstael, N.; W.M. Hanemann; I.E. Strand. 1984. "Measuring the benefits of water quality improvements using recreational demand models." Vol. II of Benefits Analysis Using Indirect or Imputed Market Methods. EPA Contract Number: CR-811043-01-0.

---

apparently no attempt was made to explore possible consequences of endogeneity.

Bockstael, N.; K.E. McConnell; and I. Strand. 1991. Recreation, in Measuring the Demand for Environmental Quality, J.B. Braden and C. Kolstad (eds.) North Holland: Elsevier Science Publishers.

Cameron, T.A.; W.D. Shaw; S. Ragland, J. Callaway, and S. Keefe. 1996. Using actual and contingent behavior data with varying time aggregation to model recreation demand. Forthcoming, *J. of Agr. and Resource Econ.*

Creel, M. and J. Loomis. 1990. Theoretical and empirical advantages of truncated count data estimators. *American J. of Agr. Econ.*, **72**: 434-441.

Ekstrand E. 1994. Economic benefits of resources used for rock climbing at Eldorado Canyon State Park, Colorado. Unpublished PhD dissertation, Department of Agricultural Economics, Colorado State University, Fort Collins, CO.

Englin, J. and J. Shonkwiler. 1995. Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation. *Rev. of Econ. and Stat.* **77**: 104-112.

Greene, W.H. 1994. Accounting for excess zeroes and sample selection in Poisson and negative binomial regression models. Draft manuscript, Dept. of Economics, New York University.

Hausman, J.; G. Leonard; D. McFadden. 1995. A utility-consistent, combined discrete choice and count data model: assessing recreational use losses due to natural resource damage. *J. of Public Econ.*, **56**: 1-30.

Hellerstein, D. 1992. The treatment of nonparticipants in travel cost analysis and other demand models. *Wat. Res. Research* **29**: 1999-2004.

Jakus, P. and W.D. Shaw. 1996. An empirical analysis of rock climber's responses to hazard warnings. Forthcoming, *Risk Analysis* (August issue).

Johnson, N. and S. Kotz. 1969. Discrete Distributions New York: John Wiley.

Morey, E.R. 1985. Characteristics, consumer surplus, and new activities. *J. of Public Economics* 26: 221-236.

Morey, E.R. 1994. What is consumer's surplus per day of use, when is it a constant independent of the number of days of use, and what does it tell us about consumer's surplus? *J. of Env. Econ. and Manage.* **27**: 257-70.

Morey, E.R.; W.D. Shaw; and R.D. Rowe. 1991. A model of recreation participation and site choice when complete trip data are unavailable. *J. of Env. Econ. and Manage.* **20**: 181-201.

Parsons, G. and M.J. Kealy. 1995. A demand theory for number of trips in a random utility model of recreation. *J. of Env. Econ. and Manage.*

Shaw, D. 1988. On site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification. *J. of Econometrics*. **39**: 211-23.

Shonkwiler, J.S. 1995. Systems of Travel Cost Demand Equations. Proceedings, 8th Interim Report for the W-133 Regional Project, Monterey, CA.

Shonkwiler, J.S. and W.D. Shaw. 1996. Hurdle count data models for recreation demand analysis. Forthcoming, *J. of Agr. and Resource Econ.*

Terza, J. and R. Wilson. 1992. Analyzing frequencies of several types of events: a mixed multinomial-poisson approach. *Rev. of Econ. and Stat.* 108-115.

Yen, J.S. and V. Adamowicz. 1994. Participation, Trip Frequency and Site Choice: A Multinomial-Poisson Hurdle model of recreation demand. *Canadian J. of Agricultural Economics*, **42**: 65-76.

Table 1
Results of Multinomial Logit Estimation[a]
N = 183 climbers

Variable Definition	Parameter Estimate (Asymptotic standard errors)
Site-specific constant term for Ragged Mountain	-0.436 (0.023)**
Implicit price divided by 100	-1.30 (0.048)**
Number of climbs in the climber's ability range divided by 100	0.274 (0.046)**
Log likelihood at convergence	-2472

[a]Estimates obtained using Gauss statistical package.

** Significant at the five percent level.

Table 2
Double Hurdle Count Data Model[a]
N = 183

Variable	Parameter Estimate (Standard Errors [b])
Participation Hurdle	
Ability (leading level)	0.282 (0.077)***
Importance of environmental education	0.181 (0.084)**
Frequency: Positive Trips Portion	
Constant term	0.066 (0.006)***
Implicit price divided by 100	-0.358 (0.046)***
Number of climbs in the climber's ability range divided by 100	1.11 (0.018)***
Log likelihood at convergence	-2103.7

[a]Estimates obtained using Gauss statistical package.

***, ** Significant at the one and five percent levels, respectively.

Table 3
Results of Joint Multinomial-Poisson/DH Model[a]
N = 183

Variables	Parameter Estimate (Standard Error)
Double Hurdle Model	
Participation Part of Model	
Constant	0.596 (0.337)*
Importance of environmental education	0.398 (0.232)*
Trip Frequency Part of Model	
Constant	1.87 (0.233)***
Ability	0.373 (0.083)***
Inclusive Value	0.337 (0.154)**
Multinomial Logit Model	
Site-specific constant term for Ragged Mountain	-3.61 (0.431)***
Implicit price divided by 100	-1.42 (0.166)***
Number of climbs in the climber's ability range divided by 100	0.549 (0.124)***
Log likelihood at convergence	-2404

[a] Estimates obtained using FIML program in Gauss.
***, **, * Significant at one, five, and ten percent levels, respectively.

Table 4
Consumer's Surplus Estimates For Percentage Reductions and Increases In Available Climbs at Mohonk Preserve Using Two Empirical Methods

	Estimation Method	
	Multinomial Logit (Site Choice Model Only)[a]	Joint Site-Choice and DH Trip Number Model[b]
Per Trip CV		
10% decrease, Mean Maximum, Minimum	\$0.02 \$0.04, \$0.002	
50% decrease, Mean Maximum, Minimum	\$0.10 \$0.18, \$0.01	
10% increase, Mean Maximum, Minimum	\$0.02 \$0.04, \$0.002	
50% increase, Mean Maximum, Minimum	\$0.11 \$0.21, \$0.01	
Annual/Seasonal CS		
10% decrease, Mean Maximum, Minimum		\$1.76 \$3.52, \$0.00
50% decrease, Mean Maximum, Minimum		\$7.85 \$16.00, \$0.001
50% increase, Mean Maximum, Minimum		\$10.35 \$20.33, \$0.02

[a]CV is the multinomial logit "per-trip" compensating variation.

[b]Seasonal E[CS] is averaged across the sample of 183 members for the increase and decrease in all available climbs at the Mohonk Preserve at a leader's ability level.

Heterogeneous Preference of Environmental Quality and Benefit Estimation in Multinomial Probit Models: A Simulation Approach

Heng Z. Chen
Stephen R. Cosslett *

May 1996

Abstract

A random parameter multinomial probit model is estimated using the simulated MLE to accommodate the varying tastes or varying perceptions of environmental quality across individuals. The expected maximum utility is also simulated for the benefit estimation. The empirical results indicate that the random parameter probit fits the data substantially better than the independent probit model which is similar to the independent logit model in both the parameter and benefit estimates. Furthermore, the benefit of improving the site quality decreases due to the existence of heterogeneous preference of the environmental site quality.

Key words: recreational site choice, random parameter and constant parameter multinomial probit models, simulated maximum likelihood estimation, valuation for environmental policy.

*The authors are Visiting Assistant Professor, Department of Agricultural Economics, Michigan State University, and Professor, Department of Economics, The Ohio State University, respectively.

Heterogeneous Preference of Environmental Quality and Benefit Estimation in Multinomial Probit Models: A Simulation Approach

The random utility model of recreational fishing is often specified as $U_j^i = p_j^i\alpha + x_j\beta + \epsilon_j^i$, where p_j^i is the trip cost to site j for individual i, x_j is the environmental quality for site j. α and β are the constant preference parameters. The stochastic term ϵ_j^i allows for the idiosyncratic taste variations across individuals, which is not observed by the researcher. If ϵ_j^i follows the type I extreme value (EV) distribution, or the generalized EV distribution, the model becomes independent, or nested multinomial logit, respectively. The logit models have been widely used to estimate the destination (or site) choice in recreation demand studies under the hypothesis of random utility maximization. See Parsons and Kealy 1992, Morey, Rowe, and Watson 1993, Kling and Herriges 1995, among others. Benefits can be estimated using the models for a measure of environmental improvement in x_j (Hanemann 1982).

However, the logit models have some undesirable properties. The well known Independence of Irrelevant Alternatives (IIA) maintained by the independent logit model restricts the pattern of substitutions across alternatives and makes the model less likely to reflect reality. Although the nested logit model using the generalized EV distribution relaxes the IIA assumption, the correlation admitted by the model is limited. More importantly, for the recreational site choice model, the environmental site quality variables x_j in the model specification are often measured as technical numbers or indices by environmental agency, which vary across sites, but not individuals. The logit models with the constant parameters α and β can be inadequate because by taking a trip to site j, the perceived site quality $z^i(x_j)$ can be differently from the indices x_j for different individuals. The utility difference between any two individuals i_1 and i_2, $U_j^{i_1} - U_j^{i_2} = (p_j^{i_1} - p_j^{i_2})\alpha + \epsilon_j^{i_1} - \epsilon_j^{i_2}$, is not a function of site quality x_j. One way to reflect the "varying perception" is to specify $z_j^i = x_j(1 + \eta^i)$, where η^i allows the perception variation for individual i. Those who are not fully informed about the techical site quality indices x can have their own local knowledge about the site quality for each site. For example, three quality indices in our data set are 1) the salmon catch rate; 2) forest coverage in percentage; and 3) a dummy variable to indicate whether the site is contaminated. When different individuals visit the same site, the salmon catch rate can be different due to weather conditions, fishing experience, local knowledge, etc.. Also, when a site is considered as contaminated using the criteria by the environmental agency, individuals might not be fully informed although there exist ways for the individuals to access the site quality information [1]. Thus, by substituting the perceived quality z_j^i for the actual quality x_j, the model becomes

$$\begin{aligned} U_j^i &= p_j^i\alpha + z_j^i\beta + \epsilon_j^i \\ &= p_j^i\alpha + x_j\beta + x_j\beta\eta^i + \epsilon_j^i \\ &\equiv p_j^i\alpha + x_j\beta + x_j\delta^i + \epsilon_j^i \end{aligned}$$

It can be seen that the constant parameter logit models are mis-specified due to the existence of the varying perceptions of site quality $x_j\delta^i$ which is not observed by the researcher.

Another way to see the restrictiveness of the logit models is that the constant parameters β's imply the constant marginal utility of site quality for all individuals. The specification does not allow for the "varying tastes" for the site quality. To reflect the varying marginal utility of the site quality across individuals, one can specify a

[1] See, for example, "Michigan Fishing Guide" which is published annually since 1983.

random parameter model with $\beta^i = \beta + \delta^i$ for the site quality, where β is the "average" taste, and δ^i is the individual specific taste variations. The heterogeneous tastes or heterogeneous marginal utility of the site quality can thus be accommodated [2]

$$\begin{aligned} U_j^i &= p_j^i \alpha + x_j \beta^i + \epsilon_j^i \\ &= p_j^i \alpha + x_j \beta + x_j \delta^i + \epsilon_j^i \end{aligned}$$

It is clear that the "varying perception" interpretation or the "varying taste" interpretation are indistinguishable. The logit models can not accommodate the complicated covariance structure of $x\delta + \epsilon$. As a result, to allow for the idiosyncratic taste variations of the unobserved utility components ϵ and the hetergeneous preference $x\delta$ of the included site quality variables, a random parameter multinomial probit (VPMNP) needs to be estimated by assuming that δ and ϵ are independent, each of them follows a multivariate normal distribution.

Although Hausman and Wise 1978 investigated transportation mode choice using a similar random parameter specification, the study was limited to investigating three alternatives (two after normalization in the model estimation). The limited computing power and econometric estimation methods at that time prevented them from evaluating the choice probabilities for the multinomial probit model when the number of choice alternatives is more than four. For recreational fishing, the number of feasible sites can be very large. Different individuals can have different feasible

[2] The trip cost parameter α can also be specified as a random variable. However, we think that this is not likely to be as important as that for site quality x_j because the trip cost p_j^i is already site specific and individual specific. Furthermore, since the utility is ordinal scaled, we can always normalize U_j^i by a preference parameter, such as α: $\frac{1}{\alpha} U_j^i = p_j^i + x_j \frac{\beta^i}{\alpha} + \frac{\epsilon_j^i}{\alpha}$. The model can be estimated once the distribution of $\{\frac{\epsilon_j^i}{\alpha}, \frac{\beta^i}{\alpha}\}$ or $\{\frac{\epsilon_j^i}{\alpha}, \frac{\delta^i}{\alpha}\}$ is specified.

sites as well. For example, there can be as many as 37 feasible sites for some individuals in our application. It is interesting to examine whether there exists any difference between the widely used independent multinomial logit model (FPMNL), the independent multinomial probit model (FPMNP), and the varying parameter (correlated) probit model (VPMNP) with the complicated correlation structure induced by either $x\delta$ or ϵ. Such comparisons were not possible until recent advances in simulation methods by McFadden 1989, Pakes and Pollard 1989, among others.

Furthermore, when the correlated VPMNP model is applied to recreational fishing, we are also interested in estimating the expected maximum utility from the model, thus, the benefits for a change of environmental quality at some sites. To this end, we have to estimate the expected maximum utility from the model before and after the policy implementation, or the inclusive values in the case of logit models (Hanemann 1982). Similar to the choice probability, there does not exist an analytic solution for the expected maximum utility. In this paper, we have to use a simulation method to estimate the expected maximum utility, thus the benefit due to a policy change, in the probit models.

This paper is organized as follows. After the VPMNP model is formulated, we discuss the simulated maximum likelihood estimation using the smooth recursive normal simulator, known as the GHK sampling method due to the independent effort by Geweke 1991, Hajivassiliou and McFadden 1990, and Keane 1990. For the benefit estimation, the expected maximum utility is simulated using the unbiased frequency simulator for the probit models. The detailed steps to simulate the choice probabilities and the expected maximum utility are listed in the appendices. By using 1983/84 survey data on Michigan anglers' recreational fishing, we estimated three models, which are then compared to assess the implications of different distribution assumptions with and without the varying tastes or varying perceptions for the site quality x. The benefit due to a policy change is estimated and compared as well. We conclude the paper by some final remarks.

Simulated Maximum Likelihood Estimation and Expected Maximum Utility

Simulated MLE of Multinomial Probit Models

To model the heterogeneous preference for the site quality across individuals, a random parameter model can be specified as

$$U_j^i = p_j^i\alpha + x_j\beta^i + \epsilon_j^i = p_j^i\alpha + x_j\beta + u_j^i \tag{1}$$

where the stochastic term is given by $u_j^i = x_j\delta^i + \epsilon_j^i$. If $\delta_k^i = 0$, the random parameter specification for the kth site quality variable is not informative [3]. If this holds for all $k = 1, ..., K$, the model degenerates to the conventional constant parameter model.

To estimate the parameters in (1), we assume that ϵ and δ are independent, each of them follows a multivariate normal distribution with $\epsilon \sim N(0, \Sigma_\epsilon)$ and $\delta \sim N(0, \Sigma_\delta)$. When $\Sigma_\delta = \text{diag}(\sigma_{\delta_1}^2, \cdots, \sigma_{\delta_K}^2)$, the covariance matrix for u is given by $\Sigma_u = \Sigma_\epsilon + \Sigma_{x\delta}$, where

$$\Sigma_\epsilon = \begin{pmatrix} \sigma_{\epsilon_{11}}^2 & \cdots & \sigma_{\epsilon_{1J}}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{\epsilon_{J1}}^2 & \cdots & \sigma_{\epsilon_{JJ}}^2 \end{pmatrix}$$

and

$$\Sigma_{x\delta} = \begin{pmatrix} \sum_k \sigma_{\delta_k}^2 x_{1k}^2 & \cdots & \sum_k \sigma_{\delta_k}^2 x_{1k}x_{Jk} \\ \vdots & \ddots & \vdots \\ \sum_k \sigma_{\delta_k}^2 x_{1k}x_{Jk} & \cdots & \sum_k \sigma_{\delta_k}^2 x_{Jk}^2 \end{pmatrix}$$

It is clear that if $\delta_k \neq 0$ for any k, the off-diagonal elements of the covariance matrix Σ_u are non-zeros. Thus, the correlations across alternatives can be introduced in two

[3]From hereafter, the superscript i is omitted to simplify the notation in this section.

ways. One is due to the varying taste or varying perception for the site quality with $\sum_k \sigma_{\delta_k}^2 x_{jk} x_{j'k} \neq 0$ for $j \neq j'$. The other is due to the correlations contained in ϵ with $\sigma_{\epsilon_{j,j'}}^2 \neq 0$ for $j' \neq j$.

To estimate the multinomial probit model, we need to evaluate the choice probability $Pr(j)$

$$Pr(j) \;=\; \int_{-\infty}^{\infty} du_j \int_{-\infty}^{(p_j-p_1)\alpha+(x_j-x_1)\beta+u_j} du_1 \cdots \int_{-\infty}^{(p_j-p_J)\alpha+(x_j-x_J)\beta+u_j} du_J \cdot f(u)$$

where $f(u)$ is a multivariate normal density function for u with mean 0 and covariance Σ_u. Unlike the multinomial logit model in which $Pr(j)$ can be expressed as ratio of the exponential functions, the difficulty in evaluating $Pr(j)$ for the multinomial probit model is the high dimension integration. To overcome this difficulty, many simulators have been introduced recently to approximate the choice probabilities through Monte Carlo simulations. For examples, the frequency method (Lerman and Manski 1981); the importance sampling method (McFadden 1989); the Stern's method; or the smooth recursive sampling method by Geweke 1991, or Hajivassiliou and McFadden 1990, or Keane 1990, among others. In this paper, we choose to simulate the choice probabilities by the smooth recursive sampling (hereafter the GHK) method since it is continuous in the parameter space. The estimation can be achieved by using standard optimization packages. Furthermore, the GHK simulator has been shown unbiased for any given number of replications R (Börsch-Supan and Hajivassiliou 1993). Based on the rooted mean squared error criterion, Hajivassiliou, McFadden, and Ruud 1992 show that the GHK simulator is unambiguously the most reliable method for simulating normal probabilities, compared to twelve other simulators considered. Börsch-Supan and Hajivassiliou 1993 also compare the GHK simulator with the frequency simulator and the Stern's simulator, and show that the GHK simulator generates substantially smaller variance than the others. For our application, the detailed steps of constructing the GHK probability simulator are provided in Appendix A.

To estimate the parameters using the simulated maximum likelihood estimation (SMLE) method, the choice probabilities in the likelihood function are replaced by the simulated probabilities. As the sample size and the number of replications in the simulation increase, maximization of the simulated likelihood yields the parameter estimates which possess the asymptotic properties of the conventional ML estimates. See Gourieroux and Monfort 1993 [4]. Statistical inference for the SMLE can also be implemented.

Expected Maximum Utility

Expected maximum utility $\bar{U}_m$ of making a choice is:

$$
\begin{aligned}
(2)\bar{U}_m &= \int_{u=-\infty}^{+\infty} \max_j (p_j\alpha + x_j\beta + u_j) f(u) du \\
&= \sum_{j=1}^{J} \int_{u=-\infty}^{+\infty} (p_j\alpha + x_j\beta + u_j) I[p_j\alpha + x_j\beta + u_j \geq p_l\alpha + x_l\beta + u_l,\ \forall\ l] f(u) du
\end{aligned}
$$

where the indicator $I(A) = 1$, if A is true, 0 otherwise. For example, when u follows the type I EV distribution for the logit models, the expected maximum utility $\bar{U}_m$ has a closed form solution $\ln[\sum_{j=1}^{J} e^{x_j\beta}] + \gamma$ and is often referred as the inclusive value in the literature (McFadden 1977), where $\gamma = 0.577215649\cdots$ is the Euler constant. However, when u follows a normal distribution with the density function $f(u)$, there do not exist analytic expressions for the expected maximum utility $\bar{U}_m$. In this paper,

[4] Other estimation methods such as the method of simulated moments (MSM), or the method of simulated scores (MSS) can also be used. Each of them shares some advantages and disadvantages which are not our focus here. For a review, see Gourieroux and Monfort 1993, Hajivassiliou, McFadden, and Ruud 1992.

we constructed an frequency simulator to estimate the expected maximum $\bar{U}_m$

$$\hat{\bar{U}}_m = \frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{J}(p_j\alpha + x_j\beta + u_j^r)I[p_j\alpha + x_j\beta + u_j^r \geq p_l\alpha + x_l\beta + u_l^r, \ \forall\ l]$$

It is clear that this simulator is unbiased since $E(\hat{\bar{U}}_m) = \bar{U}_m$. See Appendix B for the detailed steps of simulating the expected maximum utility.

Benefit Estimation

For the recreational demand studies, one of the motivations to estimate the random utility logit or probit model is to estimate the benefit for a measure of environmental site quality change. If the marginal utility of dollar $-\alpha$ remains unchanged before and after the site quality change from x^0 to x^1 for any sites, the benefit can be estimated by

$$\Delta W(x^1|_{x^0}) = \frac{\bar{U}_m(x^1) - \bar{U}_m(x^0)}{-\alpha} \tag{3}$$

where the numerator measures the difference of the expected maximum utility after and before the policy implementation. The denominator converts the utility difference into dollar unit. See Hanemann 1982 for the discussion of (3) or McFadden 1981 for the use of the expected maximum utility as the welfare function.

Data and Estimation Results

In this paper, three models are estimated using a subset of the recreation trip data contained in the report by Jones and Sung 1992. Although the details of the data set can be found in the report, we should briefly discuss the data set which consists of two parts. One is 1983/84 survey of Michigan anglers by the Michigan Department of Natural Resources. The survey questionnaires were mailed out throughout the fishing season, asking about the most recent fishing trip. From the returned survey questionnaires, 338 salmon fishing trips that last within one day to the Great Lakes (Lake Michigan, Lake Superior, Lake Huron, and Lake Erie) are selected for this study. The elementary site is defined as county with the Great Lake shoreline. There are a total of 41 Great Lake sites in Michigan that support the salmon fishing. The feasible set for each individual consists of all the sites that are within the maximum driving distance observed in the survey data set. For every feasible site, the trip cost $Cost_j^i$ is the round trip driving distance between each individual's home site and the feasible site multiplied by the AAA milage cost at 0.28 dollar per mile. Thus, $Cost_j^i$ is individual and site specific. Table I reports the summary statistics of the trip cost variable for 338 trips.

The other part of the data set is the environmental site quality variables which are also provided by the Michigan Department of Natural Resources. The summary statistics of the following site quality variables for the 41 sites can also be found in Table I.

1. A 0-1 dummy variable Aoc_j with value 1 to indicate if site j is designated as the "area of concern" for toxic contamination by the International Joint Commission, 0 otherwise. It is noticed that this index only intends to qualitatively reflect the site contamination due to, for example, Mercury, PCBs, dioxin, etc. contained in the fish body caught at the site [5]. Individuals can still fish at site j even if the site is with $Aoc_j = 1$.

2. The forest coverage of the site in percentage $Forest_j$.

3. The number of salmon caught per hour at the site j month t $Salmon_{jt}$. That the variable is site j and month t specific is because the salmon catch rate in

[5] For the details, see Michigan Fish Contaminant Monitoring Program, 1992.

the Great Lakes can change significantly over time in the fishing season (from April to October). For the details, see Jones and Sung 1992.

Using the trip cost and site quality variables, let the utility of a fishing trip to site j in month t be specified as

$$\begin{aligned} U_{jt}^{i} &= Cost_{j}^{i}\alpha + Aoc_{j}\beta_{1}^{i} + Forest_{j}\beta_{2}^{i} + Salmon_{jt}\beta_{3}^{i} + \epsilon_{j}^{i} \\ &= Cost_{j}^{i}\alpha + Aoc_{j}\beta_{1} + Forest_{j}\beta_{2} + Salmon_{jt}\beta_{3} + u_{jt}^{i} \end{aligned}$$

where the stochastic term is $u_{jt}^{i} = Aoc_{j}\delta_{1}^{i} + Forest_{j}\delta_{2}^{i} + Salmon_{jt}\delta_{3}^{i} + \epsilon_{j}^{i}$. Three models are estimated to examine the implications of different distribution assumptions for ϵ and importance of the varying taste or varying perception specification δ. Model I uses the type I EV distribution for ϵ in which ϵ_j's are independent with $\delta^i = 0$. It is an independent multinomial logit (FPMNL) model and estimated by the conventional maximum likelihood method. Table II presents the parameter estimates and t-statistics. The log likelihood value is -523.21.

Model II is an independent multinomial probit (FPMNP) model. It maintains the independent assumption for ϵ_j's, but the EV distribution is replaced by the standard normal distribution with $\Sigma_\epsilon = \mathrm{diag}(1, 1, ..., 1)$. The taste variation $\delta^i = 0$ is maintained, as in Model I. The parameter estimates using the SMLE are reported in Table III. The number of replications is 2,000. The log likelihood value actually decreases from -523.21 (Model I) to -539.89 (Model II) due to the change of distribution assumption for ϵ. This indicates that the EV distribution fits the data set better than the normal distribution, provided that the error terms are truly independent. It is noticed that to compare the FPMNL model with the FPMNP model, one has to multiply the parameter estimates of the logit model by 0.625 due to the distribution difference (Amemiya 1981). It can be seen that the estimated logit model and probit model are quite similar. For example, if we multiply the estimated marginal utility per hundred dollars ($-\alpha = 17.300$) from the FPMNL model by 0.625, the result is 10.723, fairly close to that of the FPMNP model ($-\alpha = 9.122$).

As the sample size and the number of replications increase, the parameter estimates of the SMLE possess the same asymptotic properties as that of the MLE. The GHK probability simulator is unbiased for any fixed number of replications. However, there is no guideline as to how many replications is empirically needed for our given problem. In order to verify that 2,000 replications is appropriate, the FPMNP model is also estimated with several other numbers of replications (R=10, 50, ..., 2,000). We found that as the number of replications increases, the log likelihood value gradually increases and stabilizes around -540. The parameter estimates also stabilize. When the number of replications is small, such as 10, the simulation noise can be quite severe. Table IV presents the relationship between the log likelihood value and the number of replications for the FPMNP model.

The third model (Model III) we estimated is the correlated VPMNP model in which the covariance matrix is given by $\Sigma_\epsilon + \Sigma_{x\delta}$, where $\Sigma_\epsilon = \text{diag}(1, ..., 1)$ [6]. That is, in order to compare the FPMNP with the VPMNP, we assume that the correlation across alternatives is only caused by $x\delta$ due to the heterogeneous preference of Aoc, $Forest$, and $Salmon$. Table V reports the parameter estimates for the VPMNP model with 2,000 replications [7].

The log likelihood value of the VPMNP model increases to -473.88, as compared with -539.89 for the FPMNP model due to the varying taste or varying per-

[6] A more general model with $\sigma^2_{\epsilon_{jj'}} \neq 0$ for $j \neq j'$ can be estimated if we have enough data observations to recover the $J \times (J+1)/2$ number of parameters in the covariance matrix Σ_ϵ.

[7] When the number of replications is 10, 100, 500, 1,000, the log likelihood value is -491.46, -475.43, -473.40, -474.30, respectively. The relationship is very similar to Table IV.

ception specification by adding the three parameters σ_{δ_1}, σ_{δ_2}, and σ_{δ_3}. This indicates that the independent FPMNP model does not adequately describe the data variations, thus, the parameter estimates are likely inconsistent because the independence assumption across alternatives cannot hold.

For the VPMNP model, the estimates of α, β_1, β_2, and β_3 are significant. The trip cost and the three site quality variables are useful in guiding individual's site choices. It is noticed that since the FPMNP model and the VPMNP model have different covariance structure, comparison of the parameter estimates between the two models is not as straightforward as the comparison between the independent FPMNL and FPMNP models. Furthermore, σ_{δ_1} for *Aoc* and σ_{δ_2} for *Forest* are also significant. This implies that the utility (or disutility) derived from the site quality *Aoc*, and *Forest* are different for different individuals. On average, while individuals prefer the sites without the contamination and with high forest coverage, there exists significant heterogeneity in tastes or perceptions for the site quality.

It is interesting to note that although the salmon catch rate variable is informative, (i.e. $\hat{\beta}_3$ is significant,) the heterogeneous preference of the catch rate does not exist. Difference in individual's tastes or perceptions of the catch rate is very small. Individuals do have a fairly homogeneous preference as to which Great Lake sites are good for salmon fishing. This can be because the monthly specific catch rate information of the Great Lake sites is more informative than the other two quality variables, and the information of the Great Lake sites is easily accessible, as compared to rivers or streams where this may not be the case.

We further estimated and compared the benefit due to a change in site quality for the three estimated models. The policy scenario considered in this paper is to clean up the environmental toxic contamination by setting $Aoc = 0$ for the fourteen contaminated Great Lake sites in Michigan. By using (3), the benefit of the clean up in the FPMNL model is \$1033.66 dollars for the 338 fishing trips (or \$3.06 dollars per trip) in the sample.

For the independent FPMNP model, we estimated the expected maximum before and after the policy implementation using (3) with the frequency simulator. By using 10,000 replications, the simulated benefit is \$1159.18 dollars (or \$3.43 dollars per trip). In order to ensure that the reported estimate is reliable with the 10,000 replications, we also experimented with several other different replications. For example, when $R = 100$, the estimated benefit is already stabilized around \$1160 with little fluctuation. The similarity in the benefit estimates resembles the similarity in the parameter estimates between the independent probit and logit models. For the correlated VPMNP model, the simulated benefit is \$248.56 dollars (or \$0.74 dollars per trip) for the sample using 10,000 replications. The stability is similar to the FPMNP model. The estimated benefits can be found in Table VI. It can be seen that the welfare impact of the policy change is dampened due to the taste and/or perception variations of the site quality across individuals.

Final Remarks

By comparing the FPMNP model and the VPMNP model, one can see that the varying parameter specification improves the model's goodness of fit drastically. The specification can be important because in many cases the explanatory variables, such as the site quality, is measured by a set of technical numbers that do not vary across individuals. The concern is whether the site quality indices combined with the constant preference parameters can adequately describe each individual's taste or perception of the site quality. From the estimation results, it is suggested that varying parameter specification provides a significant improvement over the constant parameter specification.

The estimation results also suggest that the varying parameter specification may be useful for some of the site quality variables included in the model, but not for the others. While there exist taste or perception variations of *Aoc* and *Forest*, the taste or perception variations of the catch rate does not exist. Individuals are sensitive to whether a site has a good salmon catch rate. The preference of the monthly specific catch rate information in the Great Lake sites appears to be uniform in our sample because the Great Lake site information may be easily accessible than the fishing sites in rivers or streams.

Furthermore, in many empirical situations, we are also interested in assessing the policy impacts using the logit or probit models. By using the frequency simulator, it is shown that the benefit estimate for removing *Aoc* in the independent FPMNL and FPMNP models are similar, which resembles the similarity in the parameter estimates between the two models. However, the estimated benefit using the correlated VPMNP model is smaller than the independent FPMNP and FPMNL models. This suggests that while the benefit estimate is not sensitive to the parametric distribution assumption *per se*, the estimate can be very sensitive to the flexibility of the parametric distribution to capture the heterogeneous preference. The mis-specification can lead to inconsistent parameter estimates and valuation of environmental policy.

Table I: Data Summary Statistics

Variables	Mean	Std.Dev.	Minimum	Maximum
Cost	11.140	13.370	0.9	75.40
Aoc	0.341	0.480	0.0	1.00
Forest	0.541	0.291	0.7	0.97
Salmon	0.036	0.044	0.0	0.22

Table II: FPMNL Model (LL = -523.21)

Parameters	Estimates	t-Statistics
α $(Cost/100)$	-17.300	-16.547
β_1 (Aoc)	-1.583	-8.916
β_2 $(Forest)$	2.532	4.775
β_3 $(Salmon)$	7.407	3.542

Table III: FPMNP Model (LL = -539.89)

Parameters	Estimates	t-Stat
α $(Cost/100)$	-9.122	-15.096
β_1 (Aoc)	-0.981	-8.656
β_2 $(Forest)$	1.556	4.639
β_3 $(Salmon)$	4.564	3.144

Table IV: Log-likelihood Value of the FPMNP (LL) and Replications (R)

R	10	50	100	200	500	1,000	1,500	2,000
LL	-578.41	-549.26	-543.06	-543.96	-543.26	-541.84	-540.88	-539.89

Table V: VPMNP Model (LL = -473.88)

Parameters	Estimates	t-Stat
α ($Cost/100$)	-16.280	-14.015
β_1 (Aoc)	-3.243	-4.274
β_2 ($Forest$)	1.437	2.422
β_3 ($Salmon$)	6.474	3.175
σ_{δ_1} (Aoc)	4.354	4.219
σ_{δ_2} ($Forest$)	4.547	5.685
σ_{δ_3} $Salmon$)	0.009	0.002

Table VI: Benefit Estimates

Models	FPMNL	FPMNP	VPMNP
ΔW	\$1033.66	\$1159.18	\$248.32
R	N/A	10,000	10,000

APPENDIX A

The simulated maximum likelihood estimation (SMLE) method involves replacing the choice probabilities $Pr(j|_{p,x,\alpha,\beta,\Sigma_u})$ in the likelihood function by simulated values $\hat{P}r(j|_{p,x,\alpha,\beta,\Sigma_u})$. The GHK simulation method generates simulated probabilities that are differentiable functions of the parameter β and Σ_u, and therefore standard optimization methods can be used to maximize the simulated likelihood with initial starting points. The steps for computing the choice probability of choosing j, $\hat{P}r(j|_{p,x,\alpha,\beta,\Sigma_u})$. using the GHK simulator in our application are as follows:

1. Draw random numbers ϵ_{ij}^{r} from the uniform distribution on [0,1], with $r = 1, \cdots, R$ (the number of replications), $i = 1, \cdots, N$ (the sample size), and $j = 1, \cdots, J_i$ (the number of alternatives in the choice set for individual i. These random numbers remain fixed during the maximization of the simulated likelihood.

2. For a specific individual i in the model (1), let the observed choice be j. (The index i is fixed in steps 2-7, and will be omitted.) Normalize the utility of choice l, U_l, by the utility of choice j, U_j,

$$U_l^* \equiv U_l - U_j = (p_l - p_j)\alpha + (x_l - x_j)\beta + u_l - u_j \equiv p_l^*\alpha + x_l^*\beta + u_l - u_j$$

Define

$$v_{(-j)} = (u_1 - u_j, \cdots, u_{j-1} - u_j, u_{j+1} - u_j, \cdots, u_J - u_j)$$

and similarly for the $J-1$ vectors $p_{(-j)}^*\alpha$ and $x_{(-j)}^*\beta$. Then, the probability of the observed choice is

$$Pr(j) = Pr(v_{(-j)} \leq -p_{(-j)}^*\alpha - x_{(-j)}^*\beta)$$

where a vector inequality means that the inequality is satisfied by each component.

3. Compute $\Sigma_{(-j)}$, the $(J{-}1) \times (J{-}1)$ covariance matrix of $v_{(-j)}$, and its Cholesky decomposition $\Sigma_{v_{(-j)}} = L'L$, where L is a lower triangular matrix. Since, $v_{(-j)} = L\eta$, where η is standard normal, $\eta \sim N(0, I)$. The choice inequalities $v_{(-j)} \leq -p^*_{(-j)}\alpha - x^*_{(-j)}\beta$ can be rewritten as

$$L\eta \leq -p^*_{(-j)}\alpha - x^*_{(-j)}\beta \tag{4}$$

The relationship between η and $v_{(-j)}$ has the following recursive form:

$$v_{(-j)} = \begin{pmatrix} L_{11}\eta_1 \\ L_{21}\eta_1 + L_{22}\eta_2 \\ L_{31}\eta_1 + L_{32}\eta_2 + L_{33}\eta_3 \\ \vdots \\ \sum_{s=1}^{J-1} L_{J-1,s}\eta_s \end{pmatrix}$$

Without loss, L can be constructed such that all the diagonal elements l_{ss} are positive.

4. Draw a value $\bar{\eta}^r_1$ from the truncated univariate standard normal distribution satisfying the first inequality of (4), i.e.,

$$\eta^r_1 \leq -(p^*_1\alpha + x^*_1\beta)/l_{11}.$$

In general, to draw η^r_k from a truncated univariate standard normal distribution such that $\eta \leq c$, one can compute

$$\bar{\eta}^r_k = \Phi^{-1}(\Phi(c)\epsilon^r_{ik})$$

where Φ is the standard normal c.d.f., and ϵ^r_{ik} is the corresponding uniform variate drawn in step 1.

5. Proceed recursively to draw $\bar{\eta}_k^r$, for $k = 2, \cdots, J-1$, from the truncated univariate standard normal distribution such that

$$(L\eta)_k \leq -p_k^* \alpha - x_k^* \beta$$

holds for all k. This constraint can be rewritten as

$$\eta_k^r \leq -(\sum_{s=1}^{k-1} l_{k,s} \bar{\eta}_s^r + p_k^* \alpha + x_k^* \beta)/l_{kk}$$

where $\bar{\eta}_1^r, \cdots, \bar{\eta}_{k-1}^r$ are the previously drawn values.

6. For the first choice inequality in (4), compute the probability

$$Q_1^r = Pr(\eta_1^r \leq -(p_1^* \alpha + x_1^* \beta)/l_{11}) = \Phi(-p_1^* \alpha - x_1^* \beta / l_{11})$$

For the remaining $k = 2, \cdots, J-1$, compute the probabilities sequentially

$$\begin{aligned} Q_k^r &= Pr(\eta_k^r \leq -(\sum_{s=1}^{k-1} l_{k,s} \bar{\eta}_s^r + p_k^* \alpha + x_k^* \beta)/l_{kk}) \\ &= \Phi(-(\sum_{s=1}^{k-1} l_{k,s} \bar{\eta}_s^r + p_k^* \alpha + x_k^* \beta)/l_{kk}) \end{aligned}$$

Thus, we have

$$\hat{P}_j^r = \Pi_{k=1}^{J-1} Q_k^r$$

7. Repeat steps 4-6 for replications $r = 1, \cdots, R$. The simulated probability of $Pr(j)$ is then

$$\hat{P}r(j) = \frac{1}{R} \sum_{r=1}^{R} \hat{P}_j^r$$

8. Repeat the procedure (steps 2 to 7) for each observation in the sample, and compute the simulated log-likelihood for the current parameters.

APPENDIX B

The steps for simulating the expected maximum utility are as follows:

1. Consider the model (1). Let the covariance matrix of an observation in the sample be Σ_u. Compute the Cholesky decomposition $\Sigma_u = LL'$, where L is the lower triangular matrix. We have $u = L\eta$, where η has an independent standard normal distribution $\eta \sim N(0, I)$.

2. For replication r, draw an vector $\bar{\eta}^r$ from the normal random variables η, thus $\bar{u}^r = L\bar{\eta}^r$.

3. Compute the utility $\bar{U}_l^r = p_l\alpha + x_l\beta + \bar{u}_l^r$ for all alternatives l. Let $\bar{U}_m^r$ be the maximum utility of all the alternatives for replication r.

4. Repeat steps 2 and 3 for replications $r = 1, \cdots R$, and take the average
$$\bar{U}_m = \frac{1}{R}\sum_{r=1}^{R}\bar{U}_m^r$$

5. Repeat steps 1 to 4 for each observation in the sample, and take the sum to estimate the total expected maximum utility for the sample.

References

Amemiya, Takeshi. "Qualitative Response Models: A Survey." *Journal of Economic Literature*, Vol. XIX (December) 1981.

Börsch-Supan, A. and V.A. Hajivassiliou. "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models." *Journal of Econometrics* Vol. 58, 1993.

Geweke, John F.. "Efficient Simulation From the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. " Computer Science and Statistics: *Proceedings of the Twenty-Third Symposium on the Interface*. Alexandria, VA: American Statistical Association, 1991.

Geweke, John F., Michael P. Keane, and David E. Runkle. "Statistical Inference in the Multinomial Multiperiod Probit Model." Federal Reserve Bank of Minneapolis Staff Report 177, 1994.

Gourieroux, Christian, and Alain Monfort. "Simulation-based Inference, A survey with special reference to panel data models." *Journal of Econometrics*, Vol. 59, 1993.

Hajivassiliou, Vassilis, and Daniel McFadden. "The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises." Cowles Foundation Discussion Paper No. 967, 1990.

Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud. "Simulation of Multivariate Normal Orthant Probabilities: Theoretical and Computational Results." Cowles Foundation Discussion Paper No.1021, 1992.

Hanemann, W. Michael. "Applied Welfare Analysis with Qualitative Response Models." Working Paper No. 241. California Agricultural Experimental Station, Giannini Foundation of Agricultural Economics, University of California, Berkeley, 1982.

Hausman, Jerry A. and D. A. Wise. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preference." *Econometrica* Vol. 46, 1978.

Jones, Carol A. and Yusen D. Sung. "Valuation of Environmental Quality at Michigan Recreational Fishing Sites: Methodological Issues and Policy Applications." July, 1992.

Keane, Michael P. "Four Essays in Empirical Macro and Labor Economics." Ph.D. Dissertation. Brown University, 1994.

Keane, Michael P. "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica*, Vol. 62(1), 1994.

Kling, Cathy L. and Joseph A. Herriges "An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization." *American Journal of Agricultural Economics*, Vol. 77(4), 1995.

Lerman, Steven R. and Charles F. Manski. "On the Use of Simulated Frequencies to Approximate Choice Probabilities, *in* Structural Analysis of Discrete Choice Data with Econometric Applications", ed. by Charles F. Manski and Daniel McFadden. The MIT Press, 1981.

McFadden, Daniel. "Modeling the Choice of Residential Location *in* Studies in Regional Science and Urban Economics", ed. by Ake Andersson and Walter Isard, 1977.

McFadden, Daniel. "Econometric Models of Probabilistic Choice, *in* Structural Analysis of Discrete Choice Data with Econometric Applications", ed. by Charles F. Manski and Daniel McFadden. The MIT Press, 1981.

McFadden, Daniel. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*, Vol. 58(5), 1989.

Michigan Fish Guide. The Michigan Department of Natural Resources, 1983 through 1995.

Michigan Fish Contaminant Monitoring Program. 1992 Annual Report, The Michigan Department of Natural Resources.

Morey, Edward R., Robert D. Rowe, and Michael Watson. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing." *American Journal of Agricultural Economics,* Vol. 75(3):578-92, August, 1993.

Pakes, A. and D. Pollard. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* Vol. 57, 1989.

Parsons, George R. and Mary Jo Kealy. "Random Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics,* Vol. 68(1), February, 1992.

Stern, S. "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models." *Econometrica,* Vol. 60, 1992.

THE PERFORMANCE OF NESTED LOGIT MODELS WHEN WELFARE ESTIMATION IS THE GOAL*

by

Joseph A. Herriges and Catherine L. Kling
Department of Economics
Iowa State University

Abstract

In this paper, we examine the performance of nested logit models in the face of model specification errors. Particular attention is placed upon the impact that these errors can have on welfare predictions. The first specification error we consider arises when nested logit is the appropriate model, but the analyst chooses the wrong nesting structure. A Monte Carlo experiment is used, together with analytical results, to examine both the sign and size of the resulting bias to welfare estimates. In addition, we explore the value of two model selection criteria in choosing among various nesting structures: (1) the likelihood dominance criteria of Pollak and Wales (1991) and (2) consistency with stochastic utility maximization as identified by Daly and Zachary (1979) and McFadden (1978). The second specification error occurs when the underlying stochastic process is not consistent with the nested logit specification. In particular, we consider the case in which the underlying stochastic terms follow a multivariate normal distribution rather than NL's generalized extreme value distribution.

May 28, 1996

*The authors would like to thank Douglas Miller and participants in the W-133 Regional Project's annual meeting for their helpful comments and suggestions. Any errors that remain are, of course, our own. This research was supported by Iowa State University's Agricultural Experiment Station, the W-133 Regional Research Project, and a grant from the U.S. Environmental Protection Agency.

INTRODUCTION

Discrete choice models have become invaluable tools in characterizing consumer selection from among finite sets of alternatives, including site selection in the recreation demand literature (Hausman, Leonard, and McFadden (1995) and Morey, Rowe, and Watson (1993)), the choice of transportation mode and travel demand (Domencich and McFadden (1975), Ben-Akiva and Lerman (1985), and Train (1986)), housing choices (Börsch-Supan, (1985, 1987)), and energy conservation (Cameron (1985)). Early applications relied upon the multinomial logit (MNL) specification, which assumes an extreme value distribution for the stochastic elements of the model. However, while this standard logit model yields convenient closed form equations for choice probabilities, it suffers from the disadvantage of imposing the independence of irrelevant alternatives (IIA) assumption. The nested multinomial logit (NL) model developed by McFadden (1978) has provided one of the more popular solutions to this problem.

In a nested logit model, groups of alternatives are identified as exhibiting common patterns of correlation. A generalized extreme value distribution is then used to parameterize the stochastic elements of these groupings. A major attraction of the NL specification is that, while it relaxes the IIA assumption imbedded in multinomial logit, it retains the characteristic of closed form choice probability equations. These gains, however, come at the cost of imposing a particular nesting structure and implied correlation pattern among the alternatives. Although it is possible to formally test alternative nesting structures (Train, Ben-Akiva, and Atherton (1989); and Kling and Thomson (1996)), most economists employing NL models are content to choose a structure and proceed without such formal tests. Thus, it is often the analyst rather than the data that determines the nesting structure and, hence, the implied pattern of correlation among alternatives. Errors in identifying the correct correlation structures may in turn yield incorrect inferences from the data.

The purpose of this paper is to examine the performance of NL models in the face of model specification errors. Particular attention is placed upon the impact that these errors can have on welfare predictions. Applications of nested logit models in resource economics are often undertaken primarily for welfare evaluation (e.g.; Bockstael, Hanemann, and Kling (1987); Bockstael, McConnell, and Strand (1989); Hausman, Leonard, and McFadden (1995); Morey, Shaw and Rowe (1991); Parsons and Kealy (1992); and Morey, Rowe, and Watson (1993)). However, little is known about the sensitivity of these evaluations to model selection.

The first specification error we consider arises when nested logit is the appropriate model, but the analyst chooses the wrong nesting structure. A Monte Carlo experiment is used, together with analytical results, to examine both the sign and size of the resulting bias to welfare estimates. In addition, we explore the value of alternative model selection criteria in choosing among various nesting structures. Two very different selection criteria are considered: the goodness-of-fit of the models as judged by the likelihood dominance criteria of Pollak and Wales (1991) and their consistency with stochastic utility maximization as identified by Daly and Zachary (1979) and McFadden (1978). In evaluating these specification criteria, we consider not only their ability to identify the "true" nesting structure, but also their accuracy in choosing the NL model that yields welfare estimates that are closest to the "true" welfare measures.

The second specification error occurs when the underlying stochastic process is not consistent with the nested logit specification. In particular, we consider the case in which the underlying stochastic terms follow a multivariate normal distribution rather than NL's generalized extreme value distribution. In this case, a multinomial probit (MNP) model would be appropriate. As in the situation when the truth is logit, we examine both the potential bias resulting from this specification error and the ability of alternative specification criteria to choose the best nesting structure.

The choice of the alternative probit specification serves two purposes. First, the MNP model is generally more flexible than its nested logit counterpart in that it does not, a priori, impose a particular correlation structure between alternatives. Until recently, it has been accepted that the computational burden of estimating probit models were prohibitive when there were more than three or four alternatives (Maddala (1983, pp. 62-63)). Thus, although MNP models are more defensible theoretically, nested logit models have generally dominated the empirical literature because of their ease of estimation. However, recent advances in econometric methods (e.g., McFadden (1989) and Börsch-Supan and Hajivassiliou (1993)) now suggest that MNP may be a viable approach in many situations, though the computational burdens remain substantially higher than with NL. By examining the biases that emerge when probit applies, but nested logit is used, we can explore the importance of these new techniques, even when the two models imply very similar correlation structures. In addition, there is a body of literature suggesting that, because their underlying distributions are similar, bivariate probit and logit models will yield similar results (Amemiya (1981) and Maddala (1983)). To our knowledge, this result has not been extended to the multivariate case. Our second specification error discussion provides the backdrop for exploring this extension, both in terms of parameter estimates and welfare predictions.

NOTATION AND BASIC MODELS

In this paper, we focus attention on the three standard models of discrete choice: MNL, NL, and MNP. Each of these models typically begins with the specification of the utility U_{ij} associated with each combination of individual i and choice alternative j and using the form

$$(1) \qquad U_{ij} = V_{ij} + \varepsilon_{ij}, \quad i = 1,\dots,N; j = 1,\dots,J,$$

where V_{ij} and ε_{ij} denote, respectively, the deterministic and stochastic portions of individual utility, N denotes the number of individuals, and J denotes the number of alternatives available to each individual. The deterministic component can be modeled as a function of both individual and alternative characteristics (X_{ij}); i.e., $V_{ij} = f(X_{ij})$, with f often restricted to be linear in the X_{ij}'s. The random components in the model (i.e., the ε_{ij}'s) are assumed to capture inter- and intra-personal variation in tastes. The basic multinomial logit model results when these variations are assumed to be i.i.d. and drawn from an extreme value distribution. The probability that an individual will choose alternative j becomes:[1]

(2) $$P_j^M = \exp(V_j) \Big/ \sum_{k=1}^{J} \exp(V_k) \; .$$

The independence of irrelevant alternatives assumption imbedded in the standard logit model manifests itself in the fact that the relative choice probabilities between any two alternatives (P_j^M / P_k^M) is independent of any other alternative and its characteristics.

The nested logit model results when the vector $\varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_K)$ is assumed to be i.i.d. across individuals and drawn from the generalized extreme value (GEV) distribution. Model specification requires that the analyst separate the J alternatives into groups of similar or correlated alternatives. For example, in a recreation demand study one might group alternative sites according to the type of recreation available (e.g., shore fishing, charter fishing, etc.). Let g(j) denote the group to which alternative j has been assigned by the analyst, J(k) indexes the first alternative within the k^{th} group (k=1, ..., K), and I(k) denotes the number of alternatives within the k^{th} group. The probability that alternative j will be selected is then given by:

(3) $$P_j^N = P_{j|g(j)} Q_{g(j)}$$

where

(4) $$P_{j|g(j)} = \exp(V_j / \theta_{g(j)}) \big/ E[g(j)]$$

denotes the conditional probability of selecting alternative j once group g(j) has been selected, with

(5) $$E[k] \equiv \sum_{j=J(k)}^{J(k)+I(k)-1} \exp(V_j / \theta_k) \, ,$$

[1] The subscript i is suppressed for the remainder of the paper for notational simplicity, except when needed for clarification.

and

$$(6) \qquad Q_k = E(k)^{\theta_k} \Big/ \sum_{g=1}^{G} E(g)^{\theta_g}$$

denotes the probability that any of the alternatives within group k will be selected. The parameter θ_k is the dissimilarity parameter for group k. An important characteristic of the nested logit model is that for any two alternatives within the same group, the relative choice probabilities are independent of any other alternatives available. Unlike the MNL model, however, this is not generally true of alternatives assigned to different groups. The NL model reduces to the MNL form when $\theta_k = 1 \; \forall \, k = 1, \ldots, J$.

Finally, the MNP model results if the vector ε is distributed i.i.d. N(0,Σ), where Σ is a J×J variance-covariance matrix between the alternatives. The probability of selecting alternative 1 then becomes

$$(7) \qquad P_1^P = \int_{-\infty}^{\infty} \int_{-\infty}^{\varepsilon_1 + V_1 - V_2} \cdots \int_{-\infty}^{\varepsilon_1 + V_1 - V_J} f(\varepsilon) d\varepsilon$$

where $f(\cdot)$ denotes the pdf associate with the N(0,Σ) distribution.

THE MONTE CARLO EXPERIMENTS

Experimental Design

In order investigate the performance of the nested logit form in predicting welfare changes, we conduct a total of two Monte Carlo experiments, one in which the underlying "true" model is nested logit and one in which the error terms are normally distributed. The simulations center around a simple four alternative model and are used to evaluate three different nesting structures: MNL, a NL model nesting alternatives (1,2) and (3,4) (termed NLA), and a NL model nesting alternatives (1,3) and (2,4) (termed NLB). By varying the underlying true distribution, we can evaluate the performance of each of these structures and the ability of model selection criteria to choose the best structure for predicting welfare measures associates with changes in the availability of these alternatives.

In order to focus attention on the error specification and nesting structure, the deterministic portion of the utility function is kept simple. The utility associated with alternative j is assumed to be proportional to the cost (C_{ij}) of obtaining that alternative. Thus, equation (1) becomes:

$$(8) \qquad U_{ij} = \beta C_{ij} + \varepsilon_{ij}, \quad i = 1,\ldots,N; j = 1,\ldots,J,$$

where $-\beta$ denotes the marginal utility of income, set at 0.1.[2] For each alternative, the price of obtaining that alternative is allowed to take on one of three values: 20, 30 and 40. With four alternatives, there are a total of 81 possible price combinations. The sample size (N) used in the simulations was set at just under 1000 (972) in order to replicate these 81 combinations exactly 12 times and insure that the vector of prices C_j were orthogonal across alternatives.

The first Monte Carlo experiment assumes that the underlying "true" model is nested logit, with alternatives (1,2) and (3,4) grouped together and $\theta_k = \theta \ \forall \ k$. Thus, the NLA model is the correct specification. Employing equations (3) - (6) and (8), the resulting choice probabilities divide up the unit interval and a uniform random number generator can then be used to simulate each individual's choice of an alternative. Finally, these choices were used to estimate via maximum likelihood the three competing nested logit models: MNL, NLA, and NLB. This process was repeated 1500 times for each of ten values of the dissimilarity coefficient θ, with θ ranging from 1.0 to 0.1 in increments of 0.1.

The second Monte Carlo experiment assumes that the underlying true distribution of preferences is normal, but that the analyst erroneously applies a logit specification. However, the structure of the variance-covariance matrix for the normal distribution was chosen so as to mimic the correlation pattern implicit in the nested logit model A. In particular, the NLA structure requires that the relative choice probabilities between alternatives 1 and 2 be independent of alternatives 3 and 4 (and vice versa). A sufficient condition for this characteristic to apply in the probit case is for the variance-covariance matrix to be block diagonal. Thus, in our second Monte Carlo experiment, we assume the ε's are distributed $N(0, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \rho & 1 \end{bmatrix} \tag{9}$$

where ρ denotes the correlation between alternatives 1 and 2, as well as the correlation between alternatives 3 and 4. A priori, we would expect the NLA model to provide a better approximation of this probit system than either NLB or MNL.

In the second Monte Carlo experiment, a multivariate normal random number generator was used to construct ε_{ij}'s and U_{ij}'s for each alternative using equation (8), which in turn yielded the individual's maximum utility choice alternative. Finally, these choices were used to estimate via maximum likelihood the three competing nested logit models: MNL, NLA, and NLB. This process was repeated 1500 times for each of ten values of ρ ranging from 0 to 0.9 in increments of 0.1.

Model Selection Criteria

In most studies that consider alternative model specifications and where welfare measurement is a primary feature, researchers choose the model that provides either the best fit to the data or that yields other desirable properties such as consistency with theory. Here, we examine both goodness-of-fit and consistency with utility theory as possible criteria for choosing among the alternative models. In

[2] Note that the deterministic portion of the indirect utility function could also be written as $V_{ij} = \alpha + \beta(Y_i - C_{ij})$, but both α and βY_i will drop out of the estimation since they do not vary across alternatives. See Hanemann (1982)

particular, we seek to determine whether any or all of the chosen criteria identify the model that provides the most accurate welfare estimates.

One of the most common questions in nested logit applications is whether or not the simpler MNL form will suffice in representing preferences. Since the MNL model is nested in either NL model A or B, it is straightforward to test the MNL against each of these two nesting structures. This provides our first model selection criteria, namely whether or not the estimated θ's are significantly different from unity. Let $NEST_k$ denote the percentage of time that θ is significantly different from one using a 95% confidence level for nesting model k (k=A,B).

While it is straightforward to distinguish between MNL and any one nested logit model, distinguishing between nesting structures is more difficult. However, the likelihood dominance criterion suggested by Pollak and Wales (1991) can provide a unique ordering of models. Conceptually, Pollak and Wales argue that if, in a non-nested testing approach, one model were to be accepted over another, it would be the one with the largest likelihood value (given adjustments for degrees of freedom). Thus, it is natural to say that the model with the highest likelihood value dominates the other. In our application, the NL models A and B have the same degrees of freedom, thus application of the likelihood dominance criterion requires a simple comparison of the optimized likelihood values. This provides our second model selection criteria. Let LD denote the percentage of time, over the 1500 repetitions, that model A likelihood dominates model B.

Finally, in addition to goodness-of-fit criteria, researchers often use consistency with utility theory as a basis for judging/selecting models. In the case of NL models, the relevant criteria for global consistency with utility maximization have been derived by Daly and Zachary (1979) and McFadden (1978) (hereafter referred to as the DZM conditions). The condition guarantees that the implied probability density function will be nonnegative. In order for this condition to be satisfied, DZM show

for additional discussion.

that the dissimilarity coefficients (i.e., θ) must lie within the unit interval. Using this as the final model selection criteria, let DZM_k denote the percentage of time, over the 1500 repetitions, that NL model k passes (i.e., fails to reject) a one-tailed test of whether θ is significantly greater than one using a 95% confidence level.

Welfare Measures

The primary purpose of our Monte Carlo experiments is to evaluate the accuracy of nested logit models in estimating welfare measures in the face of specification errors. To this end, we compute for each model the compensating variation associated with three possible policy scenarios: (1) the elimination of sites 1 and 2 from the choice set (2) the elimination of sites 1 and 3 from the choice set, and (3) the elimination of site 1 from the choice set. This first policy scenario corresponds to shutting down two correlated alternatives and an entire "nest" in our simple four alternative system. In contrast, the second policy scenario corresponds to shutting down uncorrelated alternatives (1 and 3), alternatives for which similar choices (2 and 4) remain in the choice set. Scenario 3 focuses on the elimination of a single site.

Formulas for computing the compensating variation associated with the elimination of sites are provided by Hanemann (1982) and Small and Rosen (1981) for the MNL and NL specifications and employed here to estimate the welfare impact of these policy changes. Hanemann (1982) presents formulas for the compensating variation for probit models, but these are only approximations when there are three alternatives and very difficult to compute when there are more than three alternatives. Instead, we employ a simulation approach to compute the welfare measures of interest. In particular, for the policy eliminating sites 1 and 2, we can compute the average welfare loss for the 972 individual's in our sample as

$$(10) \qquad CV_{12} = \sum_{i=1}^{N} \frac{1}{N\beta} \left\{ \left[\underset{j=1,2,3,4}{Max} \{\beta P_{ij} + \varepsilon_{ij}\} \right] - \left[\underset{j=3,4}{Max} \{\beta P_{ij} + \varepsilon_{ij}\} \right] \right\} .$$

Equation (10) is simple to interpret as the difference in utilities before and after the elimination of two of the sites divided by the marginal utility of income. For the two probit Monte Carlo experiments, the process of computing CV_{12} was repeated and averaged over 1500 draws to yield simulated welfare measures. CV_{13} and CV_{1} were similarly computed.

RESULTS

Parameter Estimates

Parameter results from the Monte Carlo experiments are summarized in Table 1, providing average parameter estimates over the 1500 repetitions for each of the 20 simulation sets, ten for each experiment. We focus first on experiment 1, in which the nested logit model A is the true model.

<u>Experiment 1: Nested Logit A</u>.

There are no surprises in terms of the estimated NLA model, where the price coefficient is consistently centered at -0.10, the true value, and the average estimated dissimilarity coefficient tracks the true θ. However, the literature provides little guidance as to what one should expect in terms of either of the misspecified models (MNL or NLB). One insight comes from the argument, provided by Train, McFadden, and Ben-Akiva (1987), that the value of the dissimilarity coefficients in nested models identifies the underlying correlation patterns among the alternatives. In particular, values of θ less than one indicate that there is greater substitution between alternatives within nests than between nests. Likewise, if θ exceeds one, there is more substitution between alternatives across nests than within nests. This expectation is confirmed by our Monte Carlo simulations. As the true θ in experiment 1 becomes smaller, alternatives 1 and 2 become more correlated. Nested logit model B erroneously separates these alternatives into different nests so that the *between* nest correlation increases as the true θ decreases. The NLB model captures this increased correlation by fitting a value of θ greater than one.

Another approach to explaining the patterns of coefficients in Table 1a is to extend arguments that have been used in the literature to compare coefficients obtained from simple logit and probit models. In particular, it is well known that logit coefficients are approximately 1.6 times their counterparts obtained using a probit specification (Amemiya (1987)). Greene (1993, p. 640) suggests that this proportionality constant comes from the fact that both models are trying to explain the marginal impacts that changes in the explanatory variables have on choice probabilities.[3] Thus, in a probit model of the choice between two alternatives, the probability of choosing alternative 1 is given by $P_1 = \Phi(\beta_p' X)$, where Φ denotes the standard normal cdf, X denotes the vector of explanatory variables, and β_p denotes the vector of coefficients in the probit model. The marginal impact on P_1 of changing the k$^{\text{th}}$ explanatory variable is then given by $\phi(\beta_p' X)\beta_{pk}$, where ϕ denotes the standard normal pdf and β_{pk} denotes the k$^{\text{th}}$ element in β_p. Similarly, for a logit model, the marginal impact of changing the k$^{\text{th}}$ explanatory variable is given by $\Lambda(\beta_l' X)[1-\Lambda(\beta_l' X)]\beta_{lk}$, where Λ denotes the logistic cdf. If these marginal impacts are to be equal at the center of the distribution (i.e., $\beta' X = 0$), then the following relationship must hold:

$$(11) \qquad \beta_{lk} = \frac{\phi(0)}{\Lambda(0)[1-\Lambda(0)]}\beta_{pk} \approx 1.6\beta_{pk} \ .$$

One can extend this line of reasoning to cases in which more than two alternatives are available. Let $P_- = (P_1, \ldots, P_{K-1})$ denote the vector of choice probabilities.[4] Consider the following matrix of marginal impacts:

[3] This paragraph draws heavily on the argument and notation of Greene (1993, p. 640).

[4] The K$^{\text{th}}$ choice probability is excluded since it provides no additional information.

$$(12)\quad \begin{bmatrix} dP_1 \\ dP_2 \\ \vdots \\ dP_{K-1} \end{bmatrix} = \begin{bmatrix} \frac{\partial P_1}{\partial x_1} & \frac{\partial P_1}{\partial x_2} & \cdots & \frac{\partial P_1}{\partial x_J} \\ \frac{\partial P_2}{\partial x_1} & \frac{\partial P_2}{\partial x_2} & \cdots & \frac{\partial P_2}{\partial x_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial P_{K-1}}{\partial x_1} & \frac{\partial P_{K-1}}{\partial x_2} & \cdots & \frac{\partial P_{K-1}}{\partial x_J} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_J \end{bmatrix}$$

where x_j denotes the j^{th} explanatory variable (j=1,...,J). In shorthand notation, this relationship becomes $dP_{\_} = \Delta dx_{\_}$, where Δ denotes the matrix of marginal probability impacts. In the case of a multinomial logit model, it is straightforward to show that $\Delta = H^{MNL} diag(\beta^{MNL})$, where

$$(13)\quad H_{ij}^{MNL} = \begin{cases} P_i^{MNL}\left(1 - P_i^{MNL}\right) & i = j \\ -P_i^{MNL} P_j^{MNL} & i \neq j \end{cases}.$$

Similarly, for the nested logit model, $\Delta = H^{NL} diag(\beta^{NL})$, where

$$(14)\quad H_{ij}^{NL} = \begin{cases} \frac{P_i^{NL}}{\theta_{g(i)}}\left\{1 - \left[(1-\theta_{g(i)})P_{i|g(i)} + \theta_{g(i)} P_i^{NL}\right]\right\} & i = j \\ \frac{-P_i^{NL}}{\theta_{g(i)}}\left[(1-\theta_{g(i)})P_{j|g(i)} + \theta_{g(i)} P_j^{NL}\right] & i \neq j, g(i) = g(j) \\ -P_i^{NL} P_j^{NL} & g(i) \neq g(j) \end{cases}.$$

Following the logic of Greene (1993), if the two models are attempting to explain the same marginal impacts in equation (12), we would expect that

$$(15)\quad H^{MNL} diag(\beta^{MNL}) \approx H^{NL} diag(\beta^{NL})$$

near the center of the distribution. The problem in using the above relationship is that it yields K-1 conditions on each parameter in the model, one for each of the probabilities affected by a marginal change in the corresponding explanatory variable. For example, in our Monte Carlo exercise, J = K-1 = 3 and $\beta_k^i = \beta_0^i \ (i = MNL, NLA)$, so that equation (15) reduces to:

$$H^{MNL}\beta_0^{MNL} \approx H^{NLA}\beta_0^{NLA} \quad (16)$$

One approach to solving for the MNL parameter β_0^{MNL} in terms of the NLA parameters is to choose β_0^{MNL} so as to minimize the sum of squared deviations between the right and left-hand sides of equation (16). As demonstrated in an appendix, available from the authors upon request, for our four alternative model this yields

$$\beta_0^{MNL} \approx \left[\frac{32+34\theta^{NLA}}{66\theta^{NLA}}\right]\beta_0^{NLA} . \quad (17)$$

Notice that the translation coefficient is decreasing in θ^{NLA}. Comparing this approximation to Monte Carlo results in Table 1a, we find that for θ^{NLA} above 0.4, the approximation is extremely good, departing from the observed β^{MNL} by less than 2 percent. However, the approximation deteriorates as θ^{NLA} approaches zero, with β^{MNL} overestimated by 60 by the time $\theta^{NLA} = 0.2$.

A similar exercise can be carried out to predict the parameter estimates for the NLB model. In particular, for the four alternative case, we obtain:

$$\theta^{NLB} \approx \left[\frac{2+6\theta^{NLA}}{8\theta^{NLA}}\right] \quad (18)$$

and

$$\beta_0^{NLB} \approx \left[\frac{96+40\theta^{NLA}}{34+102\theta^{NLA}}\right]\theta^{NLB}\beta_0^{NLA} . \quad (19)$$

The relationship in equation (18) is consistent with our expectation that θ^{NLB} will be a decreasing function of θ^{NLA} and follows closely the Monte Carlo results in Table 1a, with prediction errors of less than 4 percent for θ^{NLA} above 0.4. As was the case for the MNL model, the translation factor on β^{NLB}

increases as θ^{NLA} decreases. However, the predictions for β^{NLB} are not as precise (with prediction errors reaching roughly 50 percent for $\theta^{NLA} = 0.2$).

Experiment 2: Multinomial Probit

The second experiment was designed to change the underlying true distribution (from logit to probit) while maintaining the pattern of correlation among alternatives and, hence, the nesting structure imbedded in model A. As a result, we would expect the pattern of parameter estimates not to change significantly between experiments 1 and 2. This is the case. For the nested logit model A, the price coefficient estimates are stable, ranging from -.14 to -.12.[5] The dissimilarity coefficient starts at roughly one when the correlation between alternatives is zero and falls monotonically as this correlation increases.[6] The dissimilarity coefficient in model B again starts near one and rises as the correlation between alternatives 1 and 2 (as well as 3 and 4) increases. Finally, the price coefficients in both the MNL and NLB model follow a pattern similar to that in experiment 1.

Model Specification Criteria

Table 2 summarizes the results of the model selection criteria for each of the three experiments. Again, we begin with the first experiment, in which logit model A is the true specification. Our prior expectation is that the various selection criteria would lead us to the NLA specification and this is born out by the simulations. The first criterion tests the multinomial specification (with $\theta = 1$) as a restriction on the nested logit structure (either A or B). $NEST_A$ (in column 2) indicates the percentage of time, in the 1500 replications, that θ differs significantly from unity using a 5% critical level. As expected, when θ actually equals one, this occurs roughly 5 percent of the time. However, as the true θ departs from one,

[5] The price coefficients in experiments 1 and 2 should not be the same because of differences in the variation of the underlying GEV and normal distributions. This is analogous to the bivariate case, in which a conversion factor is needed between probit and logit results (Maddala (1983, p. 23), Amemiya (1981)). The conversion factor for the bivariate case is 1.6, whereas in our multivariate case ($\rho = 0$), a conversion factor of 1.4 would appear to hold.

$NEST_A$ increases. By the time $\theta = 0.7$, over 90 percent of the simulations reject the MNL restriction. When the wrong nesting structure is used (i.e., model NLB), the MNL restriction is rejected less frequently, though the percentage of rejections ($NEST_B$) reaches 90 percent by the time $\theta = 0.4$.

The second model selection tool is Pollak and Wales (1991) likelihood dominance criterion. Column 4 indicates the percentage of time (LD) that nested logit model A likelihood dominates model B. When $\theta = 1$, and all three models are actually correct, model A dominates model B only half of the time, as expected. However, the likelihood dominance criteria quickly picks out the correct model A as θ departs from unity. By the time $\theta = 0.7$, the LD criteria picks model A over model B 95 percent of the time.

The third model selection criterion is consistency with stochastic utility maximization. Column five provides the percentage of time (DZM_A) model A passes a test that $\theta \leq 1$ (the so-called DZM condition). Model A consistently passes this test. Model B, on the other hand, quickly begins to fail a similar test (DZM_B in column six). By the time $\theta = 0.4$, consistency with stochastic utility maximization is rejected 95 percent of the time using model B's nesting structure.

Changing to a probit specification in experiment 2, the model selection criteria continue to conform to our prior expectations, leading the analyst towards nesting structure A. The similarity in the correlation patterns between the true underlying normal distribution and the structure assumed by the NLA model is picked up both by the goodness-of-fit criteria tests ($NEST_A$ and LD) and the criterion of consistency with utility theory (DZM_A and DZM_B).

Welfare Measures

The purpose of the previous section was to evaluate the performance of prominent model selection criteria, both when the true model is nested logit and when it is probit (but the correlation pattern is

[6] As noted in Maddala (1983, p. 71), the correlation between alternatives is approximately, but not exactly, equal to

consistent with nesting structure A). In each of the three experiments, the results suggest that a researcher who has estimated and compared MNL, NLA, and NLB would be likely to choose model A, as long as the degree of correlation among the alternatives was moderate or strong (say $\theta \leq 0.7$ or $\rho \geq 0.4$). The purpose of this section is to examine whether the obvious model in terms of correlation patterns (NLA), and the one chosen by the usual selection criteria, also yields the best estimates of welfare. The welfare comparisons are made for each of the experiments considering three policy scenarios: (1) the elimination of alternatives 1 and 2, (2) the elimination of alternatives 1 and 3, and (3) the elimination of alternative 1 only.

Experiment 1: When Nested Logit Model A is the Truth

Table 3a summarizes both the mean compensating variation (CV_i) associated with eliminating alternatives set i (i = {1,2}, {1,3} and {1}) and the simulated mean prediction errors resulting from the three alternative model specifications (MNL, NLA, and NLB).[7] To begin with, we note that the resulting compensating variations are consistent with expectations. First, CV_{12} is relatively insensitive to the level of the true θ. On the other hand, as the true θ declines, both CV_{13} and CV_1 decline. These latter reductions are to be expected since, as θ declines, the remaining alternatives become better and better substitutes for the choices that are being lost and so the value of the loss is not as great.

Turning to the prediction errors, we find that the NLA model clearly outperforms the competing specifications, as expected. Except when the true $\theta = 0.1$, the fitted NLA model predictions of CV_i deviate from their true values by two percentage points or less. The poor performance of the NLA model when $\theta = 0.1$ is possibly due to the difficulty in achieving convergence in this extreme case. Less than 17

(1 - θ), so that as ρ increases, we would expect θ to decrease. In our application, $\rho \approx 1.3(1-\theta)$.

[7] Since compensating variation is a nonlinear function of the estimated parameters under each of the model specifications, the mean CV_i for each iteration of the model was computed as the average compensating variation using 200 bootstrap draws from the asymptotic distribution of the fitted parameters.

percent of the random trials converged, compared to 94 percent when $\theta = 0.2$ and 100 percent in all other cases.[8]

Neither MNL nor NLB perform as well in predicting welfare changes. For MNL, the cost of dropping an entire nest (CV_{12}) is consistently underestimated, by as much as 35%. This is to be expected, since MNL views the two alternatives being dropped (1 and 2) as comparable to the remaining sites (3 and 4), whereas the true model views the lost sites as having no close substitutes. Similarly, the NLB underestimates the welfare loss from the shutting down of sites 1 and 2, since it erroneously views sites 3 and 4 as close substitutes. It is worth noting, however, that the wrong nesting structure (NLB) outperforms no nesting structure at all (MNL).

The prediction errors are not as severe in scenario 2, when sites are dropped from each of the nests (i.e., 1 and 3). In this case, MNL overestimates the welfare loss, by as much as 24 percent. The NLB model, however, continues to outperform MNL, with prediction errors that are generally less than 5 percent. When a single site is dropped (scenario 3), the percentage prediction errors for both of the incorrect nesting structures are small, remaining in the single digits except for when θ reached 0.1.

The prediction biases found in Table 3a can be attributed to two possible sources. First, one might typically be concerned about small sample bias in the constructed welfare predictions. Both the relatively large sample size used in the current simulations (N = 972)and the precision of the NLA predictions suggest that this is not the source of bias in this case. This can be further substantiated by examining the size of the small sample bias using a second order Taylor series approximations. In particular, consider the general problem in which K alternatives are available and a set of M alternatives are being removed. The true compensating variation (CV_M) will be a function of the parameters in the site selection model (φ). In practice, the site selection parameters are not known and an estimator $CV_M(\hat{\varphi})$ is constructed based on fitted parameters $\hat{\varphi}$. An approximation to the small sample bias can be

[8] To compensate for this problem, 3000 additional trials run were for the case in which $\theta = 0.1$.

obtained by constructing a second order Taylor series expansion of $CV_M(\hat{\varphi})$ around the true φ, subtracting $CV_M(\varphi)$, and taking expectations. This yields an approximate bias of:

$$(20) \qquad Bias_M = E\left[CV_M(\hat{\varphi}) - CV_M(\varphi)\right] \approx \frac{1}{2}\sum_i \sum_j \frac{\partial^2 CV_M(\varphi)}{\partial\varphi_i \partial\varphi_j}\sigma_{ij}$$

where σ_{ij} denotes the covariance of φ_i and φ_j.

In general, the size and sign of this small sample bias will be difficult to assess, depending upon the pattern of covariances among the estimated parameters and the site selection probabilities. However, for a number of special cases of interest, the small sample bias simplifies considerably. In particular, consider the situation in which (1) the site selection probabilities are roughly equal (i.e., $P_i = P_j \;\; \forall \; i,j$), (2) the nests each contain the same number of alternatives (e.g., I alternatives per nest with J nests), and (3) the site specific utility is a function only of the cost of visiting that site (i.e., $V_j = \beta C_j$). Then it is straightforward to show that the small sample bias in estimating the loss of a complete nest, $Bias_I$, reduces to

$$(20) \qquad Bias_I \approx \left(\frac{\sigma_\beta}{\beta}\right)^2 CV_I \, .$$

That is, the small sample bias is proportional to the compensating variation, with the factor of proportionality equal to the squared coefficient of variation for β. Similarly, the small sample bias in estimating the loss of one alternative from within each nest, $Bias_J$, reduces to

$$(21) \qquad Bias_J \approx \left[\left(\frac{\sigma_\beta}{\beta}\right)^2 - \left(\frac{\sigma_{\beta\theta}}{\beta\theta}\right)\right] CV_J \, ,$$

where again, the bias is proportional to the estimated compensating variation. For our Monte Carlo simulations, these small sample biases are both small. Using equations (20) and (21), the percentage biases are of the order magnitude of 0.1% or less.[9]

Given the insignificance of the small sample bias, the prediction errors found in Table 3a can be attributed to specification biases for both the MNL and NLB models. While the simulation results indicate the direction of the biases, we can also get a handle on them analytically using the parameter translation results in equations (17) through (19). In particular, consider the bias in estimating the loss of sites 1 and 2 using the MNL model (CV_{12}^{MNL}) instead of the true NLA model (CV_{12}^{NLA}). Using equation (17), CV_{12}^{MNL} can be written in terms of the NLA parameters, so that the difference between the two welfare predictions can be written in terms of the NLA parameters alone. At the mean of the sample (where $P_i = P_j$), it can be shown that:

$$(22) \qquad CV_{12}^{MNL} - CV_{12}^{NLA} \approx \frac{16\ln(2)\left(1-\theta^{NLA}\right)}{\beta^{NLA}(16+17\theta^{NLA})}$$

Thus, as long as $0 < \theta^{NLA} < 1$ and $\beta^{NLA} < 0$, the MNL model's estimate of CV_{12} will be biased downward, as we found in our Monte Carlo simulations. Furthermore, the size of that bias is a decreasing function of both β^{NLA} and θ^{NLA}. The latter result indicates that, as the nested logit structure becomes more important (i.e., θ^{NLA} falls), the MNL prediction of CV_{12}^{MNL} will become more biased.

A similar analytical exercise yields the following estimate of the specification bias for CV_{12}^{NLB}:

$$(23) \qquad CV_{12}^{NLB} - CV_{12}^{NLA} \approx \frac{\ln(2)\left(1-\theta^{NLA}\right)}{\beta^{NLA}(1+3\theta^{NLA})}\ .$$

Again, the wrong nesting structure yields a downwardly biased estimate of CV_{12} and the degree of bias increases as θ^{NLA} declines. Interestingly, a comparison of equations (22) and (23) indicates that the MNL

[9] In fact, the procedure outlined in footnote 7 corrects for the small sample bias.

bias is larger in absolute value, so that under these circumstances the wrong nesting structure is better than no nesting structure at all.

Experiment 2: When Multinomial Probit is the Truth

Turning to experiment 2, Table 3b summarizes the true compensating variations (CV_i, i = {1,2}, {1,3} and {1}) for each value of ρ when the underlying error structure is multivariate normal, as well the simulated mean prediction errors resulting from the three alternative model specifications (MNL, NLA, and NLB). The pattern of results are similar to those found in experiment 1. That is, the NLA model typically outperforms either the MNL or NLB specifications. This is what we would have anticipated, since the NLA model mimics the underlying correlation structure imbedded in the MNP model. In addition, both of the wrong nesting structures continue to underestimate the welfare losses from shutting down sites 1 and 2 (CV_{12}) and to overestimate the welfare losses from shutting down either sites 1 and 3 jointly (CV_{13}) or site 1 alone (CV_1).

There are two changes between experiments 1 and 2 worth noting, however. First, the compensating variations are generally smaller in magnitude in experiment 2. This is not surprising, since the error terms imbedded in the logit specifications have larger variances than the standard normal distribution used in the probit model. These larger variances will magnify the welfare changes which are based on expected values of maximum welfare changes.

Second, while the NLA model outperforms the other specifications, there does appear to be a positive specification bias associated with each of the compensating variation estimates. In this case, detailed analytical expression for this bias is difficult to obtain due to the lack of both a translation factors between nested logit and MNP models and simple closed-form solutions for the probit welfare measures. However, a compelling aregument for the sign of the bias can be made by appealing the basic properties of logit and probit models.

We start by writing the following general expression for the compensating variation CV_{12}:

$$(24) \qquad CV_{12} = \frac{E\left[Max(V_1+\varepsilon_1, V_2+\varepsilon_2, V_3+\varepsilon_3, V_4+\varepsilon_4) - Max(V_3+\varepsilon_3, V_4+\varepsilon_4)\right]}{\beta}$$

Evaluated at the mean of our sample, where $V_i = V_j$, this reduces to:

$$(25) \qquad \begin{aligned} CV_{12} &= \frac{E\left[Max(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) - Max(\varepsilon_3, \varepsilon_4)\right]}{\beta} \\ &= \frac{E\left[Max(0, \varepsilon_1-\varepsilon_3, \varepsilon_1-\varepsilon_4, \varepsilon_2-\varepsilon_3, \varepsilon_2-\varepsilon_4)\right]}{\beta} \end{aligned}$$

Since the primary difference between the logit and normal distributions is the greater probability mass in the tails of the logit distribution, we would expect, *ceteris paribus*, that the extreme values in equation (25) will be generally larger for logit specifications than their probit counterparts and the corresponding welfare predictions to be larger. This is consistent with the results of the Monte Carlo simulations summarized in Table 3b.

FINAL COMMENTS AND RECOMMENDATIONS

The purpose of this study was to examine the errors in welfare estimates using nested logit models under two specification errors (inaccurate nesting structures and incorrect error distributions) and to determine whether conventional goodness-of-fit tests were useful in identifying the best model for welfare evaluation. While any Monte Carlo analysis is limited in terms of the scope of the models it can address, our finding, together with accompanying analytical results, suggest a number of conclusions.

First, specification errors in terms of nesting structure can lead to seriously biased welfare estimates. This is particularly true when the alternatives being considered for removal are close substitutes (i.e., belong the same nest). In this case, a MNL model erroneously assumes the remaining sites are comparable and understates the welfare loss. Curiously, the wrong nesting structure provides slightly superior, though still biased, estimates of the welfare loss, perhaps due to the addition degrees of freedom available in fitting the site selection probabilities.

Second, while the choice of nesting structure is important, the good news is that available model selection tools appear to perform well in choosing the correct nesting structure. In particular, the likelihood dominance criteria developed by Pollak and Wales (1991) consistently selected the true NLA structure over the NLB model once the dissimilarity coefficient became less than 0.8. The DZM conditions were less helpful in this regard, with the wrong nesting structure still passing the DZM conditions 40 percent of the time, even when the dissimilarity coefficient for the true model equaled 0.6.

Finally, it has long been known that logit and probit models yield similar characterizations of bivariate choice probabilities. While the conventional wisdom has been that these similarities would likely carry forward when the number of alternatives exceeded two, there has been little evidence in the literature to support this expectations. Our Monte Carlo simulations suggest that the nested logit model, with the nests properly specified, closely follows a multivariate probit model with a block diagonal variance covariance matrix. The welfare predictions, however, appear to be biased slightly upwards due to the heavier tails in the logit model.

Table 1. Parameter Estimates

a. Experiment 1: Nested Logit A

True θ	Price Coefficient β MNL	Price Coefficient β NLA	Price Coefficient β NLB	Dissimilarity Coefficient θ NLA	Dissimilarity Coefficient θ NLB
1.0	-.10	-.10	-.10	1.01	1.00
0.9	-.11	-.10	-.11	.91	1.05
0.8	-.11	-.10	-.12	.80	1.10
0.7	-.12	-.10	-.14	.70	1.17
0.6	-.13	-.10	-.15	.60	1.24
0.5	-.15	-.10	-.17	.50	1.31
0.4	-.16	-.10	-.19	.40	1.41
0.3	-.17	-.10	-.22	.30	1.51
0.2	-.18	-.10	-.24	.20	1.59
0.1	-.19	-.10	-.24	.11	1.63

b. Experiment 2: Multinomial Probit

ρ	Price Coefficient β MNL	Price Coefficient β NLA	Price Coefficient β NLB	Dissimilarity Coefficient θ NLA	Dissimilarity Coefficient θ NLB
0.0	-.14	-.14	-.14	0.97	0.97
0.1	-.14	-.13	-.14	0.89	1.00
0.2	-.14	-.13	-.15	0.84	1.03
0.3	-.15	-.13	-.15	0.77	1.07
0.4	-.15	-.13	-.16	0.71	1.11
0.5	-.16	-.12	-.17	0.64	1.16
0.6	-.16	-.12	-.18	0.56	1.22
0.7	-.17	-.12	-.20	0.48	1.29
0.8	-.18	-.12	-.22	0.39	1.38
0.9	-.20	-.12	-.25	0.25	1.49

Table 2. Model Selection Criteria

a. Experiment 1: Nested Logit A

True θ	$NEST_A$	$NEST_B$	LD	DZM_A	DZM_B
1.0	5	4	50	97	96
0.9	19	6	60	100	92
0.8	57	10	82	100	82
0.7	92	24	95	100	62
0.6	100	45	100	100	41
0.5	100	71	100	100	18
0.4	100	91	100	100	5
0.3	100	98	100	100	0
0.2	100	100	100	100	0
0.1	100	99	100	100	0

b. Experiment 2: Multinomial Probit

ρ	$NEST_A$	$NEST_B$	LD	DZM_A	DZM_B
0.0	8	9	52	98	98
0.1	22	5	67	100	97
0.2	47	6	81	100	93
0.3	76	6	95	100	88
0.4	96	14	99	100	78
0.5	100	25	100	100	63
0.6	100	45	100	100	39
0.7	100	68	100	100	20
0.8	100	88	100	100	6
0.9	100	98	100	100	1

Table 3. Welfare Predictions

a. Experiment 1: Nested Logit A

True θ	CV_{12}: True CV_{12}	CV_{12} Prediction Error (%): MNL	NLA	NLB	CV_{13}: True CV_{13}	CV_{13} Prediction Error (%): MNL	NLA	NLB	CV_1: True CV_1	CV_1 Prediction Error (%): MNL	NLA	NLB
1.0	7.77	1	1	0	7.77	0	0	1	3.16	0	0	0
0.9	7.78	-5	1	-4	7.22	3	0	2	2.97	1	0	1
0.8	7.79	-9	1	-8	6.69	5	0	2	2.79	2	0	3
0.7	7.80	-15	1	-12	6.19	8	0	1	2.61	3	0	4
0.6	7.82	-20	1	-16	5.73	10	0	1	2.44	3	0	4
0.5	7.84	-24	1	-20	5.31	12	0	2	2.29	4	0	5
0.4	7.87	-28	1	-24	4.95	14	0	2	2.15	4	0	6
0.3	7.89	-32	1	-27	4.65	16	0	3	2.04	5	0	7
0.2	7.93	-34	1	-29	4.41	19	-2	4	1.95	6	-2	8
0.1	7.96	-35	-2	-30	4.21	24	-44	8	1.87	10	-50	12

b. Experiment 2: Multinomial Probit

ρ	CV_{12}: True CV_{12}	CV_{12} Prediction Error (%): MNL	NLA	NLB	CV_{13}: True CV_{13}	CV_{13} Prediction Error (%): MNL	NLA	NLB	CV_1: True CV_1	CV_1 Prediction Error (%): MNL	NLA	NLB
0.0	6.00	3	5	2	6.03	2	2	4	2.36	5	5	5
0.1	6.11	0	5	0	5.88	4	2	5	2.33	5	4	5
0.2	6.17	-3	5	-2	5.73	5	2	4	2.29	5	4	5
0.3	6.25	-5	5	-4	5.58	6	1	4	2.25	4	4	5
0.4	6.32	-8	5	-6	5.43	7	1	4	2.22	4	2	5
0.5	6.42	-11	6	-9	5.26	8	1	3	2.17	4	2	5
0.6	6.49	-14	6	-11	5.07	10	1	3	2.11	4	2	5
0.7	6.55	-18	6	-14	4.88	11	0	2	2.06	4	1	5
0.8	6.64	-21	5	-17	4.67	12	0	2	2.00	3	0	5
0.9	6.72	-24	6	-20	4.44	14	-2	3	1.92	4	-1	5

REFERENCES

Amemiya, T. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19(1981):483-536.

Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Applications to Travel Demand* (Cambridge, MA: M. I. T. Press, 1985).

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., 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, A. "Hierarchical Choice Models and Efficient Sampling with Applications on the Demand for Housing." *Methods of Operations Research* (1985):175-187.

Börsch-Supan, A. *Econometric Analysis of Discrete Choice: With Applications on the Demand for Housing in the U.S. and West Germany.* Lecture Notes in Economics and Mathematical Systems (New York: Springer-Verlag, 1987).

Börsch-Supan, A., and V. A. Hajivassiliou. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics* 58 (1993):347-368.

Cameron, T.A. "A Nested Logit Model of Energy Conservation Activity by Owners of Existing Single Family Dwellings." *Review of Economics and Statistics* 67 (1985):205-211.

Daly, A., and S. Zachary. "Improved Multiple Choice Models." in David Hensher and Q. Dalvi (eds.). *Identifying and Measuring the Determinants of Mode Choice* (London: Teakfield 1979):225-227.

Domencich, T., and D. McFadden. *Urban Travel Demand: A Behavioral Analysis* (Amsterdam: North Holland, 1975).

Hanemann, W.M. "Applied Welfare Analysis with qualitative Response Models." California Agricultural Experiment Station, October, 1982.

Hausman, J.A., G.K. Leonard, and D. McFadden. "A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage." *Journal of Public Economics* 56 (1995):1-30.

Kling, C.L. and C.J. Thomson. "The Implications of Model Specification for Welfare Estimation in Nested Logit Models." *American Journal of Agricultural Economics* 78(1996):103-114.

Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press, 1983).

McFadden, D. "Modeling the Choice of Residential Location." in A Karquist et al. (eds.). *Spatial Interaction Theory and Planning Models* (Amsterdam: North Holland, 1978) 75-96.

McFadden, D. "Econometric Models of Probabilistic Choice." in C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Application*(Cambridge, MA: The M. I. T. Press, 1981), pp. 198-272.

McFadden, D. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57(1989) 995-1026.

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., W.D. 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.

Parsons, G.R., and M.J. Kealy. "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics* 68 (1992) 93-106.

Pollak, R., and T. Wales. "The Likelihood Dominance Criterion." *Journal of Econometrics* 47 (1991) 227-242.

Small, K. and S. Rosen. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49 (1981) 105-130.

Train, K. *Qualitative Choice Analysis: Theory, Econometrics, And An Application To Automobile Demand* (Cambridge, MA: M. I. T. Press, 1986).

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.

A DISCRETE CHOICE MODEL OF THE DEMAND FOR CLOSELY RELATED GOODS: AN APPLICATION TO RECREATION DECISIONS*

by

J.S. Shonkwiler and W.D. Shaw
Department of Applied Economics and Statistics/Mail Stop 204
University of Nevada
Reno, NV 89557

e-mail: jss@scs.unr.edu and wdshaw@unr.edu

ABSTRACT

Most new recreation demand (travel cost) models either do an adequate job of estimating allocation choice among competing sites or total season's trips, but a poorer job of doing both simultaneously. Recently, efforts have been made to link these two models. Two issues that arise are whether aggregation of demands provides a logically consistent price index for total season's trips and whether the linked models provide general welfare measures for environmental or site quality changes. We develop and estimate an incomplete demand system which yields the true welfare measures for price and environmental or site quality changes. The system is developed by specifying a particular quasi-indirect utility function for closely related goods which is conditional on the total quantity of trips taken in a year. The resulting price index is a linear homogeneous argument of the aggregate demand function. The application is to rock climbing destinations in the northeastern United States.

* Shaw is corresponding author. We thank George Parsons for his comments on the manuscript. An earlier version of this paper was presented at the March 1996 W-133 Regional Research Conference, Jekyll Island, Georgia. This research is partially supported by the Nevada Experiment Station.

1. INTRODUCTION

In this paper we develop and estimate an incomplete model of demand for closely related goods, with an application to outdoor recreation. Empirical economists interested in estimating demand for one or more commodities rarely have all the data necessary to estimate a *complete* system of demand equations and therefore usually analyze a narrow subset of goods of the set of all those possibly consumed. An exception is when one is concerned only with broad aggregates. Acceptable assumptions about consumer behavior vary with what objectives are of most interest and what inferences can be derived from the final model. For example, in recreation modeling, one typically wants to predict recreation trip-taking behavior under destination or site fee changes or changes in environmental quality at the destinations. Consequently, accurate and meaningful welfare measures for these changes are desired.

Outdoor recreation is an activity that has been modeled with increasing sophistication in both econometric and micro-theoretic aspects and most modern approaches derive models of recreation demand for the individual.[1] Interest in and rapid development of the models stem partially from the desire for recovery of welfare measures for environmental changes that affect destination quality, and therefore affect recreational activities. The welfare measures derived in the context of several popular recreation demand models have been scrutinized, with concerns that even the traditional welfare measures for price changes are not the "true" welfare measures one would desire. Other criticisms are that a model's welfare measures are quite difficult to estimate, or that they are very limited in what they reveal about changes that actually occur in the natural environment.

We develop and apply a model based on an incomplete demand system specification. Specifically, total demand is fixed and share equations are derived, allowing the aggregate demand price index to be

[1] While there are many models of recreation demand, some of which do not involve modeling the demand for an individual (eg. Hellerstein 1995), we focus here on models which do. A recent summary of modern approaches is found in Bockstael, Strand and McConnell (1991).

obtained from them. Our approach has distinct advantages over some other models with respect to the derivation of this aggregate demand price index and the nature of the welfare measures obtained. The microeconomic theory is laid out in Section 2 and cast in the context of recent demand literature. The econometric model is specified in Section 3. In Section 4 the data and application are discussed, and empirical results are presented. We offer conclusions in the final section of the paper.

2. MICROECONOMIC FRAMEWORK

Before describing the microeconomic framework for our model (Section 2.2), we briefly review the literature on individual recreation demand models (Section 2.1). Throughout this section we assume that the demand model of interest is derived from a constrained utility maximization problem for individual *n*. Consider two sets of goods, **Q** and **S**, with prices *p* and *r*, respectively. There are J specific goods which make up the goods in the subset **Q**, and **S** is a subset which includes all other goods. These two sets are all of the goods of interest to individual n, such that total income (Y^n) is spent on these, or:

$$p'q + r's = y \tag{1}$$

If utility for person n is generally written U(Q,S), then the usual general microeconomic conditions for optimal consumption flow simply from the constrained optimization problem, which is of course to maximize U(Q,S) subject to equation (1). Demand functions result from this and all may be well and good theoretically, but the applied economist must determine a functional form for preferences and an empirical structure for estimation of the parameters in the model, including distributions for the error terms which share the properties of the random variables. If interest is in all the goods in **Q** and **S**, a simultaneous system of equations may be implied, with symmetry and adding up restrictions needing to be checked for consistency with the microeconomic theory. However, those who want to estimate demand using empirical data rarely have access to a complete set of information on the individual. The exact quantities consumed of other goods are not known. This is a problem in virtually all applied demand analyses which use data collected from survey questionnaires. It may even be that <u>what</u> dictates the goods in a relevant set such as Q is not even known, making knowledge of one or more the prices of these goods impossible to determine. Such are the likely realities confronting the economist who wants to estimate recreation demand.

2.1 *Brief review of recent recreation demand approaches*

Particular models of recreation demand arise because of the nature of the recreational activity or because one simply inherits some existing survey data with particular consequences for how recreation trips can be dealt with. Such data are virtually certain to be quite limited, for one cannot ask about all other goods consumed by the individual in a standard recreation survey (be it mail, telephone, or in-person). In fact, one can rarely ask individuals to explain where and how often they went to every single recreation destination of potential interest over some lengthy time period. For the avid recreator, this is a very large amount of information.

Economists have been doing a better job of modeling recreation using distributions consistent with the statistical properties of recreation trips. Despite attempts to closely link empirical models to microeconomic theory (discussed below), criticisms can be still levied at any approach. Particularly problematic is the meaning of resulting welfare measures.

Recreation demand models based on formal microeconomic theory typically assume preferences are separable, allowing existence of demand equations for the recreation goods (the visits to the destinations) apart from other goods (Primont 1970). Here the budget allocated to this group of separable goods is assumed to be known, and *partial* demands are a function of the prices of the goods in the group and the quantities of nonmarket commodities (eg. environmental attributes) that relate to goods in the group. Such demands are conditional on the budget allocation to the commodities in the group. As Shonkwiler (1995) suggests, there may arise a problem of interdependence between quantities demanded and group expenditures in these conditional demand models which is exacerbated when many households have zero demands. Such zero demands are very common in recreation data collected using mail surveys of general populations, and may be common for substitute recreation sites even when the data are collected in on-site surveys.

Because the assumption of separable preferences leads to what some call partial demand systems, it also yields *partial* welfare measures. The Hicksian measures (CV and EV) are often now estimable, but assuming separability means these are partial CVs and EVs. Unfortunately, only the partial compensating variation measure has, in special cases, a known relationship to the exact *full* CV measure, as the partial equivalent variation measure provides little information about the full EV (Hanemann and Morey).

The Count Data and Random Utility Models (RUM)

The two most popular current approaches to modeling recreation demand are the count data and random utility models (RUM). Each lacks important features the other offers. Let the total number of recreation trips taken (Q^n) by the individual during some long period (a year or a season), be obtained by summing all the trips (q) taken to all the sites they visit ($Q^n = \Sigma q_j$, $j = 1, \ldots J$). To estimate Q^n a version of the count data approach is often implemented (eg. Creel and Loomis; Hellerstein 1991). Count data models are also typically used to estimate q_j for only one j^{th} recreation destination, and though multi-site count data models have been estimated as a system (Ozuna and Gomez; Shonkwiler 1995), these are not yet particularly attractive for modeling the individual's choice among many possible destinations.[2]

The RUM - usually a multinomial logit (MNL) or nested logit model (NMNL) - is especially good for modeling the destination choice among a small number of possible destinations, but typically cannot be used to model the total number of trips an individual takes during some long period of time, so destination choice is assumed conditional on the individual's Q^n. The usual RUM is consistent with a separable utility function so suffers from the problems we describe above.

[2] As Shonkwiler points out, the model developed by Ozuna and Gomez is not consistent with properties that demand systems should have and though his own approach for three sites is tractable, there may be computational difficulties in handling many recreation sites.

A version of the RUM which allows modeling the number of the individual's total season trips taken assumes that per-period destination choices are "repeated" over and over for every period in the season, yielding the aggregate demand behavior by extracting this from repeated per-period results (McFadden; Morey et al. 1993; Shaw and Ozog). Not without its critics, the repeated RUM model has been recently deemed an "implausible" theory of behavior (Feather, Hellerstein and Tomasi - FHT).

Two new attempts have been made to integrate destination choice and total seasonal trip choice. One direction taken tries to make the count data model work simultaneously for multiple sites by estimating a system of Poisson equations (Ozuna and Gomez; Shonkwiler 1995). The other approach mixes the RUM and count data approaches. It uses one or more price indices derived from the RUM destination choice model as "prices" in the aggregate demand (total trips) model (eg. Parsons and Kealy; Hausman, Leonard and McFadden; Yen and Adamowicz; Shaw and Jakus). Bockstael et al. (1984) appear to have been the first to suggest use of the inclusive value from a RUM as the price index in an aggregate demand function and Bockstael, Hanemann and Kling (BHK) apply this to data. While Hausman, Leonard and McFadden also use this index and show this is consistent with microeconomic theory, there remains debate about the correct index, because of the fact that the inclusive value leads to unidirectional changes in total participation, depending on the sign of important estimated parameters (Feather, Hellerstein and Tomasi).

Incomplete Demand Systems

Finally, another appealing approach involves a dual structure of *incomplete* demand models (LaFrance; LaFrance and Hanemann). The incomplete system is a subset of a complete one. A key assumption is that prices of goods outside of the set of goods of interest do not vary. If the models are derived via the usual utility maximization problem (assuming the budget constraint is linear), demands are positive, homogenous of degree zero in prices and total income, have a symmetric negative semidefinite slutsky matrix, and total income (Y) exceeds expenditures (M_j) for any j^{th} subset of goods. Further, when the model satisfies these four conditions, exact welfare measures (the CV and EV) for price changes of the commodities in the subset can be derived from the incomplete demand system (LaFrance). We use this approach here. Our microeconomic model proceeds as follows.

2.2 *The Microeconomic model*

We derive the recreation destination demand equations first, and then develop the aggregate demand function for all the destinations. Sticking with previous notation, assume that individual n's utility function (U) is separable in q and s so that U = U(q, f(s)), and that there exists a properly defined deflator function for s goods ρ( ), such that:

$$\rho(r) \equiv 1 \qquad \textbf{(2)}$$

Let y be total household income, z be the characteristics of the J goods in the vector q, and $\alpha = \exp(z\ \zeta)$ with ζ being the vector of parameters corresponding to z. A quasi-indirect utility function (v) for the individual can be developed which reflects deflation of y and p by ρ(r) and which is conditional on $Q = \Sigma q_j$.[3] Conditional demand functions (where demand is expressed as a function of certain quantities) date back to the work of Pollak, who recognized that if an individual's allotment of the preallocated good remains fixed, then a well-behaved conditional utility function can be specified with the preallocated good as one of the arguments. Pollak motivated conditional demand specifications by showing their applicability to the treatment of non-market goods and the effect of leisure on consumption. As a particularly relevant example, consider the implication of the assumption that the amount of leisure to be devoted to a certain recreational activity is fixed over a season. If this amount of leisure is related to the total number of recreation trips an individual takes, then it is reasonable to assume the recreation site allocation decision depends on the total number of trips taken. Additionally, this specification will link the utility formulation with the estimation technique usually applied to the allocation decision. It is well known that the

[3] The quasi-utility function is preference structure conditional on fixed **p**, but has all the properties of a utility function for Q, defined in LaFrance and in LaFrance and Hanemann.

multinomial logit estimator is based on a conditional distribution which maintains that total trips are known and are fixed.

With this in mind, we specify v to be

$$v(p,Q,y) = y - \beta^{-1} \ln\sum_{j=1}^{J} \alpha_j \exp(\beta Q p_j) \tag{3}$$

where $\beta < 0$, $\alpha_j > 0$, for all j. The beauty of this particular quasi-indirect (v) utility function is that it is quasi-convex in p and (using Roy's identity) yields ordinary demand functions of the form

$$q_j = \frac{\alpha_j \exp(\beta Q p_j)\ Q}{\sum \alpha_j \exp(\beta Q p_j)} \tag{4}$$

Note that these demand equations are unusual in that they involve **Q** as right-hand side variables. As we showed above, in conventional systems of equations derivations, demand equations are usually conditional on subgroup element expenditures adding to total subgroup expenditure (M), and the demand equations therefore are functions of total group expenditure, rather than total group quantity. The importance of this is perhaps easier to see expressing the demands above in share form:

$$\pi_j = \frac{q_j}{Q} = \frac{\alpha_j \exp(\beta Q p_j)}{\sum \alpha_j \exp(\beta Q p_j)} \tag{5}$$

where Q is fixed and known. This form allows calculation of the shadow price of Q, which has been difficult to derive in other previous modeling efforts or at least has not been clearly derived in published recreation demand work. Gorman shows that the shadow price of Q is the solution to:

$$\frac{-\partial v / \partial Q}{\partial v / \partial y} = \frac{\sum_i \alpha_i \exp(\beta Q p_i)\, p_i}{\sum_j \alpha_j \exp(\beta Q p_j)} = \sum \pi_i\, p_i = \boldsymbol{P} \tag{6}$$

This shadow price derivation avoids ad hoc derivations of the price to be included to explain aggregate demand, **Q**. With the exception of HLM who show consistency of their price index (the inclusive value from the RUM) with two-stage budgeting, this sort of formal derivation is absent in other efforts to link site choice with aggregate demand. Feather, Hellerstein and Tomasi (FHT) present an aggregate demand price index similar to that above. In their index the shares are defined differently and they do not derive their price index in a utility-theoretic manner. Note that in equation (6), changes in z enter in the definition of α, and can increase or decrease the index. This differs from the two components introduced into the aggregate demand function by Parsons and Kealy, one for the expected price, and one for the expected utility from visiting a recreation destination.

The corresponding expenditure function for the quasi-indirect utility function is

$$e(p,u) = u + \beta^{-1} \ln \sum_{j=1}^{J} \alpha_j \exp(\beta Q p_j) \tag{7}$$

and the compensated adjusted demands (those conditional on Q, with prices adjusted for this conditioning) are then

$$\pi_i^c = \frac{\alpha_i \exp(\beta Q p_i)}{\sum \alpha_j \exp(\beta Q p_j)} = \frac{\alpha_i \exp(\beta p_i^a)}{\sum \alpha_j \exp(\beta p_j^a)} \tag{8}$$

where the adjusted price of the share π_i^c is written as p_i^a because $p_i^a = Qp_i$. The difference between the adjusted demands and the unadjusted demands is important in deriving the welfare measures, as will be shown below, because of the obvious importance of getting the correct price when deriving the welfare measure. Next we discuss the aggregate demands.

Aggregate Demands

Specify the quasi-indirect utility function for the nth individual's aggregate demand as:

$$h(P,y) = -\gamma^{-1} \exp(-\gamma y) - \alpha B^{-1} \exp(BP) \tag{9}$$

where $\alpha = \exp(w\ \phi)$, with w here defined as a vector of demographic characteristics and ϕ being the corresponding parameter vector, and again y is total household income. The upper case letters represent the aggregate demand equivalents of the lower case letters used for the site-specific demand functions (specifically note that **P** is the utility-theoretic price index for Q, given explicitly in equation (6)).

When expressed in the manner we use, observed total demand can be derived as a non-negative integer and is:

$$Q = \alpha \exp(BP + \gamma y) = \exp(BP + \gamma y + w\phi) \tag{10}$$

This is the conventional way of expressing the location parameter for the Poisson model. The quasi-expenditure function associated with the quasi-indirect for the aggregate demand function above (LaFrance and Hanemann) is thus

$$e = -\gamma^{-1} \ln[-\gamma u - \gamma\alpha B^{-1} \exp(BP)] \tag{11}$$

given that $\gamma > 0$. The corresponding compensated aggregate demand function (Q^c) is

$$Q^c = \frac{\alpha\exp(BP)}{[-\gamma u - \gamma\alpha B^{-1}\exp(BP)]} \tag{12}$$

We are now in a position to derive the consumer's surplus from this aggregate demand function.

Consumer's Surplus

Once the aggregate demand function is specified with the correct aggregate price, which we have shown is **P**, then welfare measures can be constructed from it. Both HLM and FHT derive their welfare measures from a similar count model aggregate demand function (HLM use a fixed effects Poisson due to having panel data). These welfare measures will be annual, or at least seasonal, as the total trips summed across the destinations taken during the year (season) comprise Q. To show the welfare effects, first rewrite the compensated total quantity (Q^c) as

$$Q^c = \frac{\alpha\ \exp(\beta\sum \pi_j^c p_j)}{-\gamma u - \gamma\alpha\beta^{-1}\exp(\beta\sum \pi_j^c p_j)} \tag{13}$$

The welfare effects can thus be determined from the following equations

$$\frac{\partial e(P,u)}{\partial p_i} = \frac{\partial e(P,u)}{\partial P}\ \frac{\partial P}{\partial p_i} = Q^c \pi_i^c\ [1 - \beta Q^c \sum_{j=1} p_j(\pi_j^c - \delta_i)] \tag{14}$$

and

$$\frac{\partial e(P,u)}{\partial z_{ik}} = Q^c \zeta_k \pi_i^c\ \sum_{j=1} p_j(\delta_i - \pi_j^c)] \tag{15}$$

where δ is the Kronecker delta. Three welfare effects can now be shown from a compensated price change.[4] The own direct price effect is $Q^c\pi_i = q_i^c$. The own indirect price effect is $\beta q_i p_i Q^c$, and this is < 0 because the recreator substitutes away from q_i. Note also that

$$-\beta p_i^c q_i^c Q^c \sum p_j \pi_j^c \; > \; 0 \tag{16}$$

This is so because of the need to add up the shares and in doing so, the cost of substitution to alternative sites must be accounted for. Finally, equation (15) yields the third effect, that of increasing or decreasing the site characteristic (z_i) at the i^{th} site.

LaFrance and Hanemann showed that in the context of their incomplete demand system, "it is not generally possible to measure welfare changes due to nonmarket effects using demand functions" (p. 270). However, in their paper this is because the structure of their aggregate demand function leads to an unrecoverable part of the expenditure function which does not allow a testable hypothesis of the necessary and sufficient conditions for welfare measurement. Our aggregate demand function is different than theirs because ours does not include z_i directly. Instead, it enters solely through the definition of the price index (equation (6)). Because of this difference, our model does not require the same necessary and sufficient condition specified by LaFrance and Hanemann which forces them to make an assumption with no behavioral consequences (their equation (54), p. 272), and therefore concerns about the "unequivocal" welfare changes associated with changes in z_i should be lessened.

---

[4] The usual definition in research applied to environmental problems couches welfare effects in terms of the definitions of the compensating (CV) and equivalent variation (EV) Hicksian measures. For example, for a price change, the CV is often written as the difference, $e(u^0,p^1) - e(u^0,p^0)$. Duality theory can be used to show that the Hicksian demand functions are the derivatives of the expenditure function in equation (14) and (15) so our definitions merely add detail usually not apparent in the usual definition.

3. ECONOMETRIC SPECIFICATION

Our application below is to three recreation destinations.[5] The demands for each of the destinations (q_j) are modeled using the multinomial logit model, one version of the RUM. (Exact specifications are explained below, but note here that the prices for the recreation destinations are calculated using the individual's travel costs to and from the destinations, adjusted to be $p_j^a = Qp_j$). To estimate the destinations shares alone, the log likelihood has the form

$$l_{MNL} = \sum_{n=1} \sum_{i=1} q_{ni} \ln \pi_{ni} \tag{17}$$

If we specify the π_i as

$$\frac{\alpha_i \exp(\beta p_{ni}^a)}{\sum \alpha_j \exp(\beta p_{nj}^a)} \tag{18}$$

where $\alpha = \exp(z_{ni}\,\zeta)$, something akin to the conditional logit model of McFadden's results, but we of course have our adjusted prices.

For the individual's aggregate demand function we use the Poisson specification, as total trips demanded are non-negative integers

$$Q = \alpha \exp(BP + \gamma y) = \exp(BP + \gamma y + z\phi) \tag{19}$$

where P is the correct aggregate demand price, or the shadow price, as above and the Greek letters are the vectors of parameters to be estimated.

---

[5] We use a subset of sites from a data set for another study, but unlike other methods such as the multi-site Poisson system (Shonkwiler 1995), we do not see empirical difficulties arising with modeling more sites.

Estimation Method

Estimation could be accomplished in two steps, by first estimating the above using MNL with the adjusted prices and constructing the estimated π_i. Then, the estimated π^c could be used to construct P according to equation (6), and equation (19) can be estimated using a count data model with the location parameter specified as

$$\lambda_n = \exp(B\hat{P} + \gamma y_n + w_n \phi) \tag{20}$$

which in turn is

$$= \exp(B \sum \pi^c_{ni}\, p_{ni} + \gamma y_n + w_n \phi) \tag{21}$$

for the n^{th} observation.

With this two-step approach the estimated standard errors would be compromised by the fact that P is itself a function of estimated parameters. To avoid this problem, a one-step estimator of both the MNL and count data model could be specified so that λ_n is expressed as

$$\lambda_n = \exp[\frac{B \sum_i \alpha_{ni} \exp(\beta p^a_{ni}) p_{ni}}{\sum_j \alpha_{nj} \exp(\beta p^a_{nj})} + \gamma y + w_n \phi] \tag{22}$$

This yields

$$L_n = \exp(-\lambda_n) \lambda_n^Q \prod_j \frac{\pi_j^{q_{nj}}}{q_{nj}!} \tag{23}$$

and all parameters can be estimated simultaneously.

Unfortunately, neither of the above estimation methods is likely to produce consistent parameter estimates, even though both the MNL and Poisson probability models are members of the linear exponential family (Gourieroux et al.). The inconsistency stems from a violation of the regularity conditions for maximum likelihood estimation (Spanos, 12.1 and 13.3). The inconsistency is illustrated in the appendix.

We avoid the problem described in the appendix if instead of defining the adjusted prices as

$$p_i^a = Qp_i \tag{24}$$

we replace the Q with the expected trips (Q'), where Q' is conditioned by the exogenous variables in the model. This new specification unfortunately introduces a circularity in the estimation model since the price index is defined by the MNL model which requires the expected Q, while the Poisson model which generates expected Q depends in turn on the price index. However, by beginning with some initial price index, the maximum likelihood estimator is iterated both over the parameters and the expected Q.

4. THE DATA AND EMPIRICAL APPLICATION

Our specific empirical application is to rock climbing destinations in the northeastern United States. The data and empirical results are briefly described and presented in this section. Little about rock climbers and their activities is known for the obvious reason that collecting information on them is quite difficult to accomplish; of the very few recreation surveys conducted in the past that have asked about climbing, all have failed to distinguish between "rock" climbers and "mountain" climbers, leading to a tremendous variation in estimates of the actual total number of rock climbers in the United States.[6]

[6] Here we examine climbers who generally use ropes and protection devices to scale dry rock faces of cliffs. More details can be found in Jakus and Shaw.

4.1 *The Data*

The data used in the empirical application were collected in Fall 1993 using a mail survey of individuals who climb and are members of a nature preserve, the Mohonk Preserve (MP) in New York state. The MP is New York's largest private nature preserve and is located in the Catskill mountains, about 65 miles from the heavily populated New York city metropolitan area. The MP is a destination climbing area for individuals from all over the world, and is arguably the most important rock climbing area in the northeastern United States.

Of approximately 2500 members who were mailed the survey, 892 usable surveys were obtained. (As a non-profit organization the Preserve had only a limited budget which enabled a one-time mailing of the survey questionnaire along with their quarterly newsletter.) While the majority of members who returned the survey live in the states of New York and New Jersey, many live in locations all over the United States. Of the 892 questionnaires that are usable, we identified approximately 273 members as actual rock climbers.

Using the survey questionnaire, we gathered data on the total number of trips that a climber took to the Preserve in 1993, as well as the total number of trips that climbers in the sample took to a few closely related important alternative northeastern climbing areas. We consider trips to three important alternative sites in our model: Ragged Mountain (located in central Connecticut), the Adirondack Mountains (in upstate New York), and the White Mountains (near Conway, New Hampshire). These climbing destinations (as well as the MP) are similar major attractions for climbers and are located fairly near major population centers, however, Ragged Mountain differs slightly from the other three in that all of its climbing routes are rather short in length.

We recognize the obvious potential for statistical bias because our sample respondents are Preserve members. For example, one would think that the MP members would perhaps be exclusively interested in climbing at the Preserve. While the MP members took more trips to the Preserve than the other three sites, many trips are reported to the latter destinations. In addition, though members, several climbers (approximately 10 percent of the estimating sample) did not take a climbing trip to any climbing destination in 1993, including the MP.

Because of our sample, we focus our empirical investigation to modeling the demand for the three alternative climbing areas mentioned above rather than on the MP. We make no attempt to make inferences about a more general climbing population. Though no data is available to verify this exactly, we note that along with the MP, our three sites likely constitute the majority of rock climbing activity in the Northeastern United States.[7]

4.2 Specification, Estimated Parameters and Consumer's Surplus

Specification

To estimate the system of demand equations for the three destinations, we need to specify the specific elements of the vectors of explanatory variables. For the MNL model, we include the price (travel cost to each site), a site characteristic for the Adirondacks and the White Mountains, and a site constant variable for the Ragged Mountain site. The site characteristic measures the number of available routes at each climbing destination, by degree of technical difficulty of the routes.[8] There is little previous modeling of rock climbing demand on which to base a measure of the climbing site characteristic - see Jakus and Shaw; Shaw and Jakus). For the total trips Poisson model, we specify the model to include the price index, income, the age of the respondent, an intercept or constant term, and a dummy variable for gender for that intercept.

---

[7] We have likely omitted only one closely related destination climbing area of potential importance, which is located in Bar Habor, Maine, which is so far away from the other three that we feel justified in omitting this from the analysis. (Except for from Boston, Massachusetts, Bar Harbor is not easily accessible to those who live in major population centers.) Finally, there are several small climbing areas scattered throughout the northeast, but they are much different in character than the four sites considered here.

[8] Ragged Mountain is qualitatively different in the nature of the routes it offers, and thus we use the site constant term to capture the effects on demand other than its price.

Estimated Parameters

The likelihood function was maximized using a nonlinear optimization procedure written for the GAUSS statistical package and the results are reported in Table 1.[9] As can be seen there, the price and characteristics parameters are all significantly different from zero except for the age variable in the total trips model. Price, or travel cost, has the expected negative influence on the probability of choosing a destination, and our site characteristic is also positive and significant for the two sites. The RM constant term, which may proxy for a site characteristic there, is negative and significant.

The significant price index in the aggregate demand model, unlike the HLM price index which has a positive sign, has the unconfusing negative sign that is what we expect in a demand model. The count data aggregate demand model also indicates that older people take fewer total trips and that gender does not appear to be a discriminating factor in total trips taken.

Estimated Consumer's Surplus

As we stated earlier, we make no attempt to make inferences to a general population from our sample of MP members, nor do we estimate welfare measures including the MP as one of the climbing destinations. Nevertheless, we think it illustrative to estimate and report the consumer's surplus from the model for Ragged Mountain, Adirondacks, and the White Mountains climbing areas. First, we estimate the system CV, or essentially the annual WTP rather than do without all of these three sites. For the three climbing areas, the CV is \$449.20 (1993 dollars). We also derive the welfare measure in equation (14) for an increase in the price of each site of one dollar, which might reflect a site fee increase. With this price increase, expenditures for RM, the Adirondacks, and White Mountains increase by \$1.15, about \$0.71, and \$0.49 respectively.

---

[9] Unlike some multi-site methods (eg. Shonkwiler 1995), our approach does not require evaluation of complicated multiple integrals, so we do not envision difficulties in application to many sites other than disbelief in use of the conditional MNL which has the extreme value distribution underlying it.

Finally, we derive the welfare measure in equation (15) for a small increase in the site characteristic (1 available climb) at each site. Expenditures at RM, Adirondacks and White Mountains change by -$0.155, $0.05, and $0.10, respectively. FHT point out that the HLM inclusive value index forces a particular change in demand for changes in recreation destination quality. As is seen here by the expenditure decrease at RM, our index allows site characteristic increases to lead to smaller expenditures at one site. However, RM is the site which has the separate site characteristic term set to zero, replaced with the dummy variable, so in essence we are adding a route with the same qualities as a route at the Adirondacks and the White Mountains to the RM site. We interpret the negative expenditure to indicate that climbers would, holding utility constant, need to spend less for a visit to RM: an amount equal to the income an individual would have to give up to be indifferent under conditions where they could get one route like at the other two areas and the initial (zero similar routes) conditions.

Two unpublished rock climbing studies yield per trip values for one climbing site of about $40 to $48 (Ekstrand) and $70 to $90 (Shaw and Jakus). Both are recreation demand models, but each model is quite different than the one developed here. In both of the other studies CS measures are derived, but for different price changes and destinations. Neither study considers simultaneous elimination of three destinations. Still, if one could make inferences for each of our destinations separately, a value of $100 WTP per trip is close to the value obtained by Shaw and Jakus. Shaw and Jakus also estimate the CV for a change in the number of available climbs at the Mohonk Preserve and obtain, as we do, very small (less than one dollar) WTP estimates.

5. SUMMARY AND CONCLUSIONS

We have developed and estimated a model based on an incomplete demand system which is consistent with the theory of utility maximization, and allows derivation of the price of the aggregate demand function. Aggregating across goods that are closely related yields a believable link between demands for each item in a group and the total demand for them. What we have shown in this paper is that summing across recreation trips to a group of similar sites is meaningful in the same way summing cans of brand name soda yields a meaningful total of "cans of soda." Our approach isn't recommended for goods that are not closely related because the aggregate demand price may mean very little. A clear derivation of this aggregate demand price in joint recreation models has not been shown before. While theirs may be consistent with two stage budgeting, our index has more desirable flexible properties than does the one used by HLM, which is the inclusive value from the jointly estimated RUM.

Our application here is to modeling the demand for rock climbing, which has only been addressed in unpublished literature. The welfare measures we derive and report are exact, and theoretically correct in that they have none of the bias that plagues the welfare measures derived in a partial demand system. No recreation demand model can be estimated without its share of theoretical and empirical problems. However, our model is tractable for many sites, and because our environmental changes are translated into a utility-theoretic price index, we believe we can derive correct welfare measures for these changes. We add that our welfare measures are flexible in that an increase (decrease) in site quality may either increase or decrease the compensating change in expenditure. This flexibility is shown to exist even though we do not specify two separate components to, using Parsons and Kealy's words, "pass through" to the aggregate demand function.

APPENDIX

To illustrate the inconsistency using the methods we describe, we rewrite equation (22) by suppressing the observational index (n), using ω to represent the exponential function of $\gamma y_n + w_n \phi$ and explicitly defining the adjusted prices, which results in the equation:

$$\lambda = \omega \exp\left[\frac{B\sum \alpha_i p_i e^{\beta Q p_i}}{\sum_j \alpha_j e^{\beta Q p_j}}\right] \tag{25}$$

Because λ is the location parameter for the Poisson process, regularity conditions require that there be no dependence between λ and the range of y. Note however, that the form of λ changes between two regimes, defined as $Q = 0$, and $Q > 0$. Denote each regime corresponding to λ and λ^0, and λ^+, respectively. Further, let

$$\lambda^0 = \omega \exp\left[\frac{B\sum \alpha_i p_i}{\sum \alpha_j}\right] \tag{26}$$

and

$$\lambda^+ = \omega \exp\left[\frac{B\sum \alpha_i p_i e^{\beta Q p_i}}{\sum \alpha_j e^{\beta Q p_j}}\right] \tag{27}$$

As long as β and B are < 0 and not all of the p_j are equal then λ^+ is greater than λ^0, with the magnitude of the difference varying systematically with increasing Q, ceteris paribus.

REFERENCES

Bockstael, N.; I. Strand; K.E. McConnell. Recreation. A chapter in Measuring the demand for environmental quality. J.B. Braden and C.D. Kolstad (eds) North Holland: Elsevier Science Publishers. (1991)

Bockstael, N.; W.M. Hanemann; C.L. Kling. Modeling recreational demand in a multiple site framework. *Water Resour. Res.* 23: 951-60. (1987)

Bockstael, N.; W.M. Hanemann; I.E. Strand. "Measuring the benefits of water quality improvements using recreational demand models." Vol. II of Benefits Analysis Using Indirect or Imputed Market Methods. EPA Contract CR-811043-01-0 (1984).

Creel, M. and J. Loomis. Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *Amer. J. of Agr. Econ.* 72: 434-41. (1990)

Ekstrand, E. Economic benefits of resources used for rock climbing, Eldorado State Canyon State Park, Colorado. Unpublished PhD dissertation. Colorado State University, Fort Collins, CO. (1994)

Feather, P.; D. Hellerstein; T. Tomasi. A discrete-count model of recreational demand. *J. of Environ. Econ. and Manage.* (**29**): 214-227. (1995)

Gorman, W. Tricks with utility functions. Chapter 11 in Essays in Economic Analysis, M.J. Artis and A.R. Nobay (eds.) Cambridge: Cambridge U. Press. (1976)

Gourieroux, C.; A. Monfort; A. Trognon. "Pseudo maximum likelihood methods: theory." *Econometrica* 52: 681-700 (1984).

Hanemann, W.M. and E.R. Morey. Separability, partial demand systems, and consumer's surplus measures. *J. of Environ. Econ. and Manage.* 22: 241-58. (1992)

Hausman, J.; G. Leonard and D. McFadden. A utility-consistent, combined discrete choice and count data model: assessing recreational use losses due to natural resource damage. *J. of Public Economics* 56: 1-30. (1995)

Hellerstein, D. Using count data models in travel cost analysis with aggregate data. *Amer. J. of Agr. Econ.* 73: 860-66. (1991)

Hellerstein, D. Welfare estimation using aggregate and individual-observation models: a comparison using monte carlo techniques. *Amer. J. of Agr. Econ.* 77:620-29 (1995)

Jakus, P. and W.D. Shaw. An empirical analysis of rock climber's response to hazard warnings. Forthcoming, *Risk Analysis*. (1996)

LaFrance, J.T. Incomplete demand systems and semilogarithmic demand models. *Australian J. of Agr. Econ.* 34: 118-131. (1990)

LaFrance, J.T. and W.M. Hanemann. The dual structure of incomplete demand systems. *Amer. J. of Agr. Econ.* (1989)

McFadden, D. Modeling the choice of residential location, in Spatial Interaction Theory and Residential Location. A. Karlquist et al. (eds) North Holland: Amsterdam. (1978)

Morey, E.R.; M. Watson, R.D. Rowe. A repeated nested-logit model of Atlantic salmon fishing. *Amer. J. of Agr. Econ.* 75: 20: 181-201. (1993)

Ozuna, T. and I.A. Gomez. Estimating a system of recreation demand functions using a seemingly unrelated Poisson regression approach. *Rev. of Econ. and Stat.* 76: 356-60. (1994)

Palmgren, J. The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika* 68 563-6. (1981)

Parsons, G.R. and M. Kealy A demand theory for number of trips in a random utility model of recreation. *J. of Environ. Econ. and Manage.* 29: 357-367. (1996)

Pollak, R.A. Conditional demand functions and consumption theory. *Quarterly J. of Economics* 83: 60-78. (1969).

Primont, D. 1970. Functional structure and economic decision making. Unpublished PhD dissertation, Univ. of California, Santa Barbara.

Shaw, W.D. and P. Jakus. Travel cost models of the demand for rock climbing. Unpublished draft manuscript, Department of Applied Economics and Statistics, University of Nevada, Reno. (1996)

Shaw, W.D. and M. Ozog. Application of a nested logit model to modeling trip length decisions. Unpublished draft manuscript, Department of Applied Economics and Statistics, University of Nevada, Reno. (1996)

Shonkwiler, J.S. Systems of travel cost models of recreation demand. Eighth Interim report for the W-133 Western Regional Research Conference. (October) (1995)

Spanos, A. Statistical foundations of econometric modeling. Cambridge U. Press: New York (1986).

Terza, J. and S. Wilson. Analyzing frequencies of several types of events: a mixed multinomial-poisson approach. *Rev. of Econ. and Stat.* 108-115. (1992)

Yen, S. and V. Adamowicz. Participation, trip frequency and site choice: a multinomial-poisson hurdle model of recreation demand. *Canadian J. of Agr. Econ.* 42: 65-78 (1994)

Table 1: Three Destinations Model (Estimated Parameters)
N = 178

Variable Name/Definition	Parameters (Standard Errors)
Multinomial Logit Model	
Price/10	-0.073 (0.0058)***
Site Characteristic[1]	0.004 (0.0019)**
Ragged Mountain Constant Term	-3.000 (0.015)***
Count Data Model	
Price Index	-0.0052 (0.0003)***
Income (in 1000's)	0.0011 (0.0012)
Age	-0.0030 (0.0022)
Intercept Constant Term	1.8600 (0.0152)***
Male Gender Dummy	-0.0060 (0.0154)
Log likelihood at convergence	-377

[1] Site characteristic is set to zero for Ragged Mountain.

***, ** indicates significance at the one, five percent levels, respectively.

CONCEPTUAL ISSUES IN USING CONJOINT ANALYSIS TO EVALUATE ECOTOURISM DEMAND - A CASE STUDY FROM BAHIA, BRAZIL

Thomas Holmes[1], Keith Alger[2], Chris Zinkhan[3], and Evan Mercer[1]

ABSTRACT

This paper uses conjoint analysis to evaluate the demand for and the value of ecotourism attributes in a threatened forest ecosystem in northeastern Brazil. Adaptive Conjoint Analysis computerized interviews were conducted with 215 tourists visiting the region. An ordinal interpretation of the rating scale was used and marginal utilities were estimated using ordered probit. The results indicated that: (1) a proxy variable representing degree of respondent involvement in the experimental task can be predicted from socio-economic and behavioral variables, (2) increasing predicted respondent involvement obviates the cognitive error of confusing commodity price with quality, and (3) predicted respondent involvement can be used to calibrate marginal value estimates. We conclude that degree of respondent involvement may act as a summary statistic for the dynamic properties of preference search in conjoint analysis exercises and that such a summary statistic is useful for identifying non-homogenous preferences in cross-sectional data.

[1] Southern Research Station, U.S. Forest Service, Research Triangle Park, North Carolina
[2] Institute for Socio-Environmental Studies, Ilhéus, Bahia, Brazil
[3] Campbell University, Buies Creek, North Carolina

INTRODUCTION

Environmental valuation methods can be utilized to evaluate conservation and development options for natural resource sectors in developing countries. Although the contingent valuation method (CVM) has become very popular for evaluating environmental values in developed economies, few examples exist on the use of CVM in developing economies. The CVM has been used to evaluate the demand for rural water supply in Nigeria (Whittington et al. 1992) and the value to rural household in Madagascar of the loss of access to tropical rain forests (Shyamsundar and Kramer 1993). The CVM and travel cost models have also been used to evaluate tourist demand for forest conservation in Madagascar (Mercer, Kramer and Sharma 1995).

Conjoint analysis methods developed for conducting marketing research are becoming popular for conducting environmental valuation studies in developed economies. This paper seeks to explore the application of this method for estimating resource conservation values in a developing economy. In particular, we use conjoint analysis to estimate marginal values for forest conservation and forest-based activities in a threatened forest ecosystem in northeastern Brazil. We are not aware of any other published studies that use conjoint analysis to estimate the value of environmental protection in developing countries.

THE ATLANTIC COASTAL FOREST IN SOUTHERN BAHIA, BRAZIL

The Atlantic Coastal Forest of Brazil (Mata Atlântica) is one of the most diverse and threatened tropical forest ecosystems in the world. The region around the Una Biological Reserve in southern Bahia (northeastern Brazil) is under a particularly severe threat of deforestation due to the collapse in world cocoa prices that has forced many farmers to cut their forests to pay expenses. The forests in this region contain very high levels of endemism and biological diversity. For example, these forests contain the only remaining native habitat of endangered primates such as the golden-headed lion tamarin and the yellow-breasted capuchin monkey. Further, a recent forest inventory found a world record number of tree species in a single hectare near Ilhéus, Bahia (Thomas and Carvalho 1993).

Ecotourism is an economic activity that may provide economic opportunities to private forest owners and help conserve forests in this region. Currently, most visitors to the Ilhéus region of southern Bahia come to visit the beaches and international visits to the coastal areas in this region of Brazil are increasing. The Inter-American Development Bank views tourism as an important economic development industry for this region and is investing significant resources to improve the tourist infrastructure. Forest conservation may play an important but unrecognized role in enhancing the tourism value of this region by providing esthetically pleasing landscapes and opportunities for forest-based recreation. Further, forest conservation may help stabilize soils and protect water quality in the region.

To assist the conservation planning efforts in this region, we developed a conjoint analysis instrument to provide information about forest protection values and potential forest attractions. The study results indicated that respondents engaged in varying levels of involvement with the survey instrument and that the degree of involvement influenced economic value estimates. We found that people with a low level of involvement displayed anomalous behavior, confusing commodity price with quality, and that this behavior was obviated by people with a high level of involvement. Also, marginal value estimates for forest conservation and nature attractions were sensitive to the degree of respondent involvement. Our results are consistent with the Whittington et al. (1992) study which indicated that CVM WTP values were significantly influenced by the time spent thinking about responses.

In the next section of the paper, we present a traditional rating scale conjoint model that is modified to include degree of involvement as a determinant of respondent behavior. We then provide a description of our survey instrument and experimental setting. This is followed by our results and, finally, our conclusions and implications for future research.

RATING SCALE CONJOINT AND RESPONDENT INVOLVEMENT

The traditional rating scale conjoint model decomposes individual preferences into systematic and random components:

$$V^{ij}(Q^{j}) = v^{i}(Q^{j}) + \epsilon^{ij}$$

where $V^{ij}(Q^j)$ is the true but unobservable utility of commodity *j* to individual *i*, $v^i(Q^j)$ is the systematic component of utility and ϵ^{ij} is a random error term with mean zero. Letting r represent individual "i's" rating of commodity *j* and *q* represent a vector of attributes for commodity j, a linear preference function can be specified:

$$r^{j} = \alpha + b_1 q_1^{j} + \ldots + b_k q_k^{j} + b_p p^{j} + e^{j}$$

where *p* is the price of commodity *j*. Equating the total differential of *r* to zero allows the marginal rate of substitution between attributes *m* and *n* to be computed as b_m/b_n and the marginal value of attribute m is b_m/b_p where b_p is the marginal utility of income. Typically, commodity ratings are regressed on commodity attributes and price to estimate the *b* vector of parameters.

In estimating this model, previous researchers have noticed variation in intra-individual mean ratings across the sample (Mackenzie 1993; Roe, Boyle and Tiesl 1996). These centering points or anchors have been simply viewed as sources of statistical noise in the estimation process. To increase estimation efficiency, Mackenzie included mean respondent ratings as an explanatory variable[1] and Roe et al. used a rating difference measure. Rather than viewing mean ratings simply as anchors to be treated as nuisance parameters, we argue that mean ratings may in fact be good proxies for degree of involvement by the respondent in the rating exercise. We hypothesize that greater personal involvement

[1] We note that it is likely that mean ratings are contemporaneously correlated with the equation error. Consequently, we suggest the use of an instrumental variable approach.

in the rating exercise is associated with task importance and consequently the investment of greater cognitive effort by respondents. In general, we expect that greater respondent effort yields more reliable responses.

This argument is based on recent research reported in the social psychology literature (Latané and Nowak 1994). Based on the general conception that attitudes can be viewed as processes and not just as points on a linear scale, the dynamic properties of attitudes become important. Latané and Nowak (LN) dispute the traditional view beginning with Thurstone (1931) that attitudes can be represented as points on a continuum and present a more modern view that attitudes act like categories. LN develop their argument by recalling Zeeman's (1977) proposition that attitude change may be described in the language of catastrophe theory. Zeeman's basic hypothesis was that involvement is a "splitting" factor, that is, the more involved people are the more strongly they are to maintain their attitude in the presence of new information and the less likely they are to be neutral. Following this logic, LN hypothesize that as involvement increases, people will tend to have either very positive or very negative attitudes. LN go on to present experimental evidence that the mean and variance of attitudinal scores increase along with respondent involvement.

Conjoint analysis experiments typically utilize a series of iterative ratings reported by individuals for various commodity bundles. Consequently, the dynamic properties of the experimental design may be important[2]. Given the cognitive effort required to respond in a thoughtful and consistent manner, it appears reasonable that people may vary in the degree of cognitive effort they invest in the exercise and that people who are more involved (or who view the exercise as more important) will invest greater effort in providing responses. In turn, this greater effort may manifest in terms of consistency, rationality, or reliability of responses.

[2] The dynamic properties of rating responses in an ACA conjoint analysis experiment are investigated by Johnson, Desvouges, Fries and Wood (1995).

Consequently, we propose to test the following hypotheses. First, the mean of intra-individual ratings can be explained by a set of personal characteristics that can be plausibly associated with personal involvement with the experimental task. Second, predicted involvement with the task (using a set of personal characteristics as explanatory variables) influences inter-individual ratings. In particular, we hypothesize that predicted involvement both shifts the entire rating scale and influences marginal valuations as well. The model we propose can be written as:

$$r^{j*} = \alpha + (b_1 + b_1^*\gamma)q_1 + \ldots + (b_k + b_k^*\gamma)q_k + (b_p + b_p^*\gamma)q_p + b_{k+1}\gamma + \epsilon$$

where b_m* is the "involvement" adjustment in parameter b_m, and γ is the predicted level of individual involvement. If predicted involvement is not significant, this model collapses to the traditional conjoint rating scale model.

ADAPTIVE CONJOINT ANALYSIS

For this experiment, we elicited conjoint responses using the Adaptive Conjoint Analysis (ACA) program provided by Sawtooth Software[3]. Respondent characteristics were elicited using Sawtooth Software's Ci3 computerized interview software. The ACA procedure uses a pairwise comparison of commodity profiles, with one profile appearing on the left of the screen and one profile on the right. Respondents are asked to indicate which profile they prefer by supplying a numerical rating between 1 (strongly prefer left profile) and 9 (strongly prefer right profile). A response of 5 indicates indifference between the two profiles displayed. Respondents are requested to supply preference ratings for a series of paired commodity profiles. Because informational efficiency is expected to be greatest for paired comparisons with similar utility, ACA selects profile pairs based on predicted respondent utility. That is,

---

[3] To gain perspective regarding ACA's effectiveness for predicting consumers' nature tour selections, we compared ACA to two other conjoint methods on a convenience sample of graduate students. Results are presented in the Appendix.

ACA attempts to quickly move to points of respondent indifference. Consequently, high intra-individual mean ratings suggest respondent resistance to the indifference-seeking strategy.

ACA utilizes OLS regressions to estimate part-worths of each attribute level. However, it appears that ACA responses are better interpreted as ordinal utility differences rather than cardinal utility values. That is, because the respondent directly evaluates the difference in utility between profiles, regression analysis should be based on differences in attribute levels:

$$
\begin{aligned}
dV_{ij} = \alpha &+ (b_1 + b_1^*\gamma)(q_1^a - q_1^b) + \ldots + (b_k + b_k^*\gamma)(q_k^a - q_k^b) \\
&+ (b_p + b_p^*\gamma)(q_p^a - q_p^b) + b_{k+1}\gamma + e
\end{aligned}
$$

where $(q_m^a - q_m^b)$ is the difference in levels for attribute m[4]. Because responses are ordinal, this equation is estimated using an ordered probit algorithm.[5]

The predicted level of individual involvement, γ, was estimated using an instrumental variable technique. First, mean values for intra-individual ratings were computed. Second, mean values were regressed on a set of exogenous explanatory variables using OLS. Finally, model parameter estimates were used to predict values for the instrumental variable γ.

EXPERIMENTAL SETTING AND DESIGN

The experiment was conducted in the region in and around Ilhéus, Bahia, Brazil - a popular tourist destination particularly for beach related recreation. Computerized intercept interviews were conducted at the beach, in local lodgings, and at local nature attractions. Of the 215 interviews completed, 200 respondents were Brazilian (interviews were conducted in Portuguese). The remainder of the interviews were conducted in English.

---

[4] We note that differences in attribute levels are computed for continuous variables only. Dummy variables for discrete attribute levels are not differenced.

[5] We used the ordered probit algorithm provided by Limdep software.

The interviews were conducted in two parts. First, as part of the Ci3 interview, people were asked to provide socio-economic-demographic information about themselves and their family and their itinerary for their current trip. Then they were asked to participate in the conjoint (ACA) interview. This section was introduced by asking respondents to consider the kind of tourism features they would want for a visit to southern Bahia. The ACA interview proceeded by introducing attributes and attribute levels. Respondents were asked to eliminate any level that was unacceptable and to indicate the importance of attribute levels. Based on this preliminary information, ACA proceeds to the pairwise comparison of profiles. Finally, respondents were asked to indicate the likelihood that they would purchase specific tourism packages composed of the various attribute levels. This concluded the interviews.

EMPIRICAL RESULTS

Descriptive statistics for the variables used in the analysis are shown in Table 1. Respondents were relatively young (mean = 36.7 years), well-educated (75% had some college education), and had above average incomes (R$2272 per month[6]). Most respondents were visiting the area primarily for beach recreation (42%), followed by nature tourism (36%), visiting friends (5%), cultural tourism (2%) and shopping (1%). Business and other reasons accounted for the remainder of visits.

Results of the OLS regression of mean ratings on a set of explanatory variables are shown in Table 2. As can be seen, the overall explanatory power of the model is good (adjusted $R^2 = 0.292$) and most of the explanatory variables are significant at the 0.01 level or higher. These results in general confirm our first hypothesis that mean ratings are a proxy for individual involvement in the experiment. For example, mean ratings are lower for people with lower incomes, people who are visiting the region to go to the beach or go shopping, and people who spend less time in the ACA interview. In contrast, mean

[6] Monetary units are Brazilian Reis. At the time of the survey, 1 R$ = US$1.12.

ratings are higher for people with higher incomes, people who are visiting the region to go on nature tours or cultural visits, and who spend more time in the ACA interview.

The model represented in Table 2 was used to predict individual involvement γ. The instrumental variable γ was entered in the ratings equation both as an autonomous variable and as an interaction variable with attribute levels.

Results of the ratings equation estimation are shown in Table 3. Several of the involvement interaction parameters are significant at the 0.05 level or higher and the involvement instrumental variable γ is significant at greater than the 0.01 level. Of particular importance we note that the involvement interaction parameter estimate on *entrance_fee* corrects an apparent anomaly or cognitive error committed by respondents. That is, the uncalibrated parameter estimate on *entrance_fee* (0.066) is positive and significant at the 0.01 level, suggesting that respondents confuse the price of nature attractions with quality (i.e. utility increases with the entrance fee). However, noting that the involvement interaction parameter on *entrance_fee* was negative and significant at the 0.01 level., our results indicate that this anomaly is obviated by increasing respondent involvement. For example, the parameter estimate on *entrance_fee* calibrated for average respondent involvement was nearly zero (0.005) and calibrated for one standard deviation greater than predicted mean involvement was negative (-0.011) as would be predicted by economic theory[7]. Finally, we note that the marginal utility of money computed from the *daily_ expenditure* variable (-0.011) is virtually the same as the marginal utility of money computed from the *entrance_fee* variable adjusted for one standard deviation greater than average involvement. Degree of respondent involvement appears to improve the consistency and reliability of imputed values.

Using -0.011 as the estimated marginal utility of money, marginal values associated with forest protection and nature attractions can be computed. The results reported in Table 3 indicate that the

---

[7] The computations are (-0.029*mean γ) + 0.066 = 0.005, where mean γ = 2.12, and (-0.029*(mean γ + s.d.)) + 0.066 = -0.011, where s.d. = 0.49.

marginal value of *forest_protection* decreases as involvement in the experiment increases[8]. For example, the marginal value of a 1% change in the amount of remaining forest cover was estimated to be $1.95 from the uncalibrated parameter, but estimated to be $0.89 from the parameter calibrated for average task importance. These values can be interpreted as the loss in value per adult visiting the region associated with loss in forest cover. Presumably, losses would be incurred both from shorter visits and from loss of visitors to the region as a result of deforestation.

The *nature_park1* variable represented "a nature park located in a small forest where visitors can see many tall trees as well as birds and free-ranging golden headed lion tamarins". The uncalibrated parameter estimate on the *nature_park1* variable was negative and significant at greater than the 0.01 level. This result indicated the counterintuitive result that the presence of a nature park in the region would decrease respondent utility. However, the parameter estimate for the variable interacting γ with *nature_park1* was positive and significant at greater than the 0.01 level. The *nature_park1* parameter estimate calibrated for the average nature tourist to the region is positive (0.0188). Dividing through by the marginal utility of money, the marginal value of a nature park to current nature tourists was estimated to be $1.71 per adult. We note that this amount is similar in magnitude to the estimated marginal value of forest protection.

Nature attraction variables are embedded in the next higher level. Consequently, the *nature_park2* variable was described as *nature_park1* plus "a walkway constructed in the forest canopy". The parameter estimate for this variable was significant at the 0.12 level. The implicit value of *nature_park2* was computed to be $12.35 per adult. The *nature_park3* variable was described as

---

[8] We note that previous research has shown that environmental values computed by conjoint analysis generally exceed values computed by contingent valuation (e.g., see Magat, Viscusi and Huber 1988). Our results indicate that environmental values computed by conjoint analysis are sensitive to the degree of respondent involvement. In particular, our results are consistent with Whittington et al. (1992) who demonstrated that CVM WTP for an environmental good decreased with increasing time spent thinking about the valuation problem.

nature_park2 plus a "botanical garden on a cocoa farm". The parameter estimate for this variable was significant at greater than the 0.01 level. The implicit value of *nature_park3* was computed to be $53.37 per adult[9].

CONCLUSIONS AND IMPLICATIONS

The results reported here indicate that the degree of involvement with a conjoint analysis experiment can be predicted from socio-economic-demographic characteristics and behavioral variables. We found that involvement in the conjoint exercise (measured by mean ratings) was significantly related with income, education, reason for current trip, and time spent performing the conjoint exercise. Marginal values computed from ratings data were sensitive to the predicted involvement proxy variable. Notably, people with a high degree of predicted involvement did not confuse commodity price with quality whereas people with a low degree of predicted involvement did make this cognitive error. Also, marginal values associated with forest protection and nature attractions were sensitive to the predicted degree of involvement. These results suggest that degree of respondent involvement may act as a summary statistic for the dynamic properties of preference search induced by conjoint analysis. Further, our results are consistent with the Whittington et al. (1992) CVM study that showed WTP decreases with increasing amounts of time to consider the valuation problem. Further research should be undertaken to test the generalizability of our respondent involvement model in other experimental contexts.

---

[9] The involvement variable γ was not significant in interactions with *nature_park2* or *nature_park3* and was therefore dropped from analysis.

LITERATURE CITED

Johnson, F. R., W. H. Desvouges, E. E. Fries, L. L. Wood. 1995. Conjoint analysis of individual and aggregate environmental preferences. *TER Technical Working Paper No. T-9502*, Research Triangle Park, NC.

Latané, B. and A. Nowak. 1994. Attitudes as catastrophes: from dimensions to categories with increasing involvement, in (R. R. Vallacher and A. Nowak, eds.) *Dynamical Systems in Social Psychology*. Academic Press, Inc. New York. p. 219-249.

Mackenzie, J. 1993. A comparison of contingent preference models. *American Journal of Agricultural Economics* 75:593-503.

Magat, W. A., W. K. Viscusi, and J. Huber. 1988. Paired comparison and contingent valuation approaches to morbidity risk valuation. *Journal of Environmental Economics and Management* 15:395-411.

Mercer, E., R. Kramer, and N. Sharma. 1995. Rainf forest tourism - estimating the benefits of tourism development in a new national park in Madagascar. *Journal of Forest Economics* 1: 239-269.

Roe, B., K. J. Boyle, and M. F. Teisl. Deriving estimates of compensating variation from conjoint data. Unpublished manuscript.

Shyamsundar, P. and R. A. Kramer. 1993. Does contingent valuation work in non-market economies? *Western Regional Research Publication, W-133 Benefits & Costs in Natural Resource Planning, Sixth Interim Report*. p. 64-78

Thomas, W. and A. M. Carvalho. 1993. Projeto Mata Atlântica Nordeste: Estudo Fitosociologico de Serra Grande, Uruçucu, Bahia, Brasil. *Paper presented at the XLIV Congresso Nacional de Botanica*, São Luis, Maranhão, Brasil.

Thurstone, L. 1931. The measurement of attitudes. Journal of Abnormal and Social Psychology 26:249-269.

Whittington, D. V., V. K. Smith, A. Okorafor, A. Okore, J. L. Liu, and A. McPhail. 1992. Giving respondents time to think in contingent valuation studies: a developing country application. *Journal of Environmental Economics and Management* 22:205-225.

Zeeman, E. C. 1977. *Catastrophe Theory*. Addison-Wesley, Reading, MA.

Table 1. Variable descriptions

Variable name	Description	Value[1]
forest_protect	Amount of forest remaining	0%, 50%, 100%
entrance_fee	Fee per nature attraction (R$)[2]	5, 10, 20, 25
daily_expend	Food and lodging cost per adult (R$)	25, 50, 100, 150, 200
congest1	Rare traffic congestion	0, 1 dummy
congest2	Occassional traffic congestion	0, 1 dummy
congest3	Frequent traffic congestion	0, 1 dummy
lodge1	Camping facilities	0, 1 dummy
lodge2	Simple lodging (no air cond.)	0, 1 dummy
lodge3	Nice lodging (w/air cond.)	0, 1 dummy
lodge4	Luxury lodging	0, 1 dummy
lodge5	Exclusive resort	0, 1 dummy
nature_park0	Present situation	0, 1 dummy
nature_park1	View flora and fauna in forest	0, 1 dummy
nature_park2	Nature_park1 + canopy walk	0, 1 dummy
nature_park3	Nature_park2 + botanical garden	0, 1 dummy
γ	Predicted involvement = mean rating	2.12
income	Monthly income (R$)	2272.1
educ	Dummy variable 1 = has some college, 0 otherwise	0.753
age	Respondent age, years	36.717
nature	Purpose of trip = nature tourism	0.356
beach	Purpose of trip = beach tourism	0.421
culture	Purpose of trip = cultural tourism	0.023
friends	Purpose of trip = visit friends	0.047
shopping	Purpose of trip = shopping	0.012
time	Time spent in ACA exercise in minutes	9.462

[1] Values for attributes are attribute levels. Other values are mean values.
[2] Monetary units are Brazilian Reis. At the time of the survey, 1 Reis = $1.12.

Table 2. OLS regression results predicting mean intra-individual ratings

Variable	Parameter estimate	t-ratio
constant	1.702	16.381
income (1,000)	0.057	2.649
educ	-0.197	-3.750
age	0.001	0.504
nature	0.196	2.821
beach	-0.192	-2.802
culture	1.133	7.074
friends	0.005	0.040
shopping	-0.886	-4.323
time	0.041	14.172

N = 1212
Adj $R^2 = 0.292$

Table 3. Ordered probit estimates of marginal utility

Variable	Parameter estimate	t-ratio
constant	-0.996	-5.593
Δforest_protect	0.021	3.931
Δforest_protect*γ	-0.005	-2.214
Δentrance_fee	0.066	2.655
Δentrance_fee*γ	-0.029	-2.563
Δdaily_expend	-0.011	-2.454
Δdaily_expend*γ	0.003	1.469
congest1	0.124	1.380
congest2	-0.257	-3.001
congest3	-0.045	-0.187
lodge1	0.030	0.270
lodge2	-0.051	-0.485
lodge3	0.257	2.494
lodge4	-0.064	-0.466
lodge5	0.491	2.484
nature_park0	0.024	0.079
nature_park1	-2.131	-3.335
nature_park1*γ	0.804	2.919
nature_park2	0.136	1.550
nature_park3	0.586	5.853
γ^	0.982	12.431
μ1	0.947	19.251
μ2	1.573	27.993
μ3	2.067	34.081

N = 1211
$\chi2 = 303.019$

Appendix: Predictive validity of three conjoint analysis methods

In order to gain perspective regarding ACA's effectiveness for predicting consumers' nature tour selections, the authors replicated the Brazilian study with a convenience sample of 77 U.S. graduate business (MBA) students. In addtion, the authors elicited preference data using two alternative techniques - a compositional method and traditional full profile conjoint analysis - and gathered preference data associated with hold-out nature tour alternatives. With respect to hit rate, both ACA and traditional conjoint analysis outperformed a random choice model. As a further test of validity, the mean Kendall tau's between the directly ranked hold-out alternatives and the rank orders estimated with each of the three respective preference elicitation techniques yielded tau's not significantly different than zero (at the 0.05 level). In contrast to the hit-rate test, the latter results suggest a lack of validity.

Appendix Table 1. Percent of correct first-choice predictions (using utility data to predict hold-out choices)

	Compositional	Full-profile	ACA
All respondents (n=77)	27.3%	40.3%**	40.3%**
Group evaluating 4 attributes (n=21)	47.6*	33.3	57.1**
Group evaluating 5 attributes (n=27)	18.5	40.7*	25.9
Group evaluating 6 attributes (n=29)	20.7	44.8*	41.4*

* Significantly greater than a random choice model at the 0.05 level.
** Significantly greater than a random choice model at the 0.01 model.

Appendix Table 2. Mean Kendall tau: each respondent's direct versus utility-based rankings

	Compositional	Full-profile	ACA
All respondents (n=77)	0.13	0.35	0.31
Group evaluating 4 attributes (n=21)	0.30	0.37	0.47
Group evaluating 5 attributes (n=27)	-0.04	0.41	0.27
Group evaluating 6 attributes (n=29)	0.17	0.27	0.23

Note: None are significantly different from zero at the 0.05 level (with testing applied to normalized data, zero mean and unit standard deviation).

FLEET TURNOVER AND OLD CAR SCRAP POLICIES*

by

Anna Alberini
University of Colorado
Boulder, CO

Winston Harrington
Resources for the Future
Washington, DC

and

Virginia McConnell
University of Maryland, Baltimore Co.
Baltimore, MD
and
Resources for the Future
Washington, DC

May 1996

Abstract.

This paper examines the factors affecting owners' valuations of their old vehicles using a unique longitudinal dataset. Willingness to accept for the vehicle is well predicted by mileage and condition of the car. Our estimated model of vehicle value is used as an input into a simulation model of a 1,000-car fleet representative of California's fleet. Other inputs into the simulation models are the estimated distributions of emissions in the fleet, and two equations that link emissions reductions to the cost of repairs. The simulation model is used to examine the role of scrap policies alone and combined with other policies for reducing emissions, such as current I/M programs and proposed emissions fees, and the welfare implications of combining such programs. The model incorporates both technical and behavioral relationships, and assumes that of all possible options (repairing the car, scrapping the vehicle or turning it in to an old car scrap program, paying the emissions fee without repairing the vehicle) the owner chooses the one with the least cost. We find that old car scrap program may increase net welfare under a regulatory program like I/M in practice today, but that a stand alone scrap program is unlikely to provide very much in the way of emission reductions.

* Research for this project was partially supported by a grant from the EPA. We would like to thank Don Stedman, Doug Lawson, Gary Bishop and Paul Durkin for providing most of the data for the simulation model, and Jean Hanson for outstanding research assistance.

1. INTRODUCTION

Despite dramatic reductions in "new car" emissions standards over the past 20 years, vehicle emissions continue to be a major source of urban air pollution in the U.S. The reasons for this are complex and numerous[1], but the main elements appear to be the increase in the number of vehicles and the deterioration in performance of emissions control equipment as vehicles age. Because new cars are cleaner than older cars, sometimes dramatically so[2], policies that encourage turnover of the fleet or early scrappage of older vehicles have at least the promise of significant emission reductions.

Not only have newer model year vehicles become less polluting since the early 1970s, there is new evidence that stricter warranty regulations on the post 1991 vehicles have resulted in vehicles whose emissions equipment is less likely to deteriorate over time than before. To the extent this is true, the policy focus will shift even more in the direction of reducing old car emissions and increasing the pace of fleet turnover.

Probably the most politically attractive of policies that encourage scrappage of older vehicles are programs that pay a bounty (usually \$500 to \$1,000) to owners of older vehicles if they turn their vehicle in to be scrapped. These programs are voluntary and appear to politicians and the public to be low-cost, especially when public tax monies are not used to finance them. Most scrap programs so far have in fact been privately financed, usually by companies seeking emissions offsets or relief from other regulations. All have been one-off affairs, primarily designed to demonstrate the feasibility of the idea. In 1994, however, California enacted as part of its State Implementation Plan (SIP) a program scrap 75,000 older vehicles per year for ten years, using as a scrappage inducement a bounty of up to \$1,000 per vehicle. Vehicle scrappage has thus become an important component of California's plan to meet air quality targets.

---

[1] The formation of ambient ozone is very complex and it is not clear how much of current levels comes from mobile sources and how much is from other man-made or natural sources. Also there is substantial regional variation in what causes ozone formation. See National Research Council (1991).

[2] For example, we found in an earlier study (Alberini, Edelstein, Harrington and McConnell (1993) that some pre-1980 model year vehicles had emissions of hydrocarbons (HC) as high as 25 grams per mile, while a new 1992 model year car on the road at the same time would have HC emissions less than .4 grams per mile.

Unfortunately, accelerated scrappage programs suffer from some severe limitations. While several studies have shown them to be at least moderately cost-effective, their emissions reduction potential is small unless the scrap bounty is very large, which substantially reduces their cost effectiveness (Alberini, Harrington and McConnell 1995). A large-scale scrappage program might also have large price effects in used car markets. In addition, observers are skeptical of the perverse incentives that might accompany a long-term scrappage program (Alberini et al, 1994).

The contrast with the main policy directed at in-use vehicle emissions, namely inspection and maintenance (I/M), could hardly be more complete. I/M policies have at least the potential (as yet unrealized) for very large emission reductions, but to make them effective will very likely arouse intense hostility from motorists. First implemented in the early 1980s, I/M programs produced at best modest results, and in 1990 Congress directed the EPA to develop regulations for an "Enhanced" I/M program that would correct the presumed deficiencies of the existing state programs. The Enhanced I/M regulations have proved to be extremely controversial, and last year the Agency essentially withdrew its insistence that the states adopt some of their more onerous features. I/M is now in limbo. An effective I/M program would also encourage vehicle retirement, obviously, by raising the relative cost of driving older vehicles. But such a program will not be implemented until something is done to overcome public opposition.

Given the shortcomings of pure scrappage on the one hand and pure I/M on the other, some observers have suggested combining the two. A motorist facing a $450 repair bill to get an inspection certificate for his 1974 Dodge Dart is not likely to be a supporter of I/M. A scrap bounty of $500 might mollify him. However, little is known about the properties of such hybrid programs, and in fact there is not much empirical data on motorist scrap decisions in the first place, let alone how those decisions might operate in an environment containing both I/M and scrappage inducements.

In this paper we model the decision to scrap a car at the household level and estimate its determinants using longitudinal data on the actual decisions of owners of older vehicles. In the second part of the paper, we use the empirical results to incorporate the scrappage decision into a model of fleet emissions and examine the emissions reductions and welfare implications of various policies directed at mobile source emissions reductions. These policies include pure scrappage, scrappage combined with I/M and scrappage combined with vehicle emissions fees. These emissions fees are based on the results of the vehicle emissions tests. Vehicle emissions fee policies have not been implemented in any jurisdiction to our knowledge, but our results suggest that they are a plausible alternative to either scrappage programs or existing I/M.[3]

This paper is organized as follows. We first derive a model of vehicle ownership and scrappage, then estimate an equation for the value of old cars using data from the Delaware Retirement Program surveys in section 2. We then describe the simulation model of the fleet emissions, and build in a scrappage program (alone or incorporated within an I/M or emissions fee program) in Section 3. Finally, we show the welfare results of different scrap policies in section 4. Section 5 concludes.

2. A MODEL OF THE DECISION TO SCRAP A VEHICLE

Despite the importance of fleet turnover for policies to reduce vehicle emissions, surprisingly little is known about the behavior that underlies car ownership decisions, particularly the decision to scrap. There is little evidence about which cars are scrapped and why, or about the distribution of vehicle prices or values as cars age. Statistics are available only for *average* vehicle values and *average* vehicle scrap rates by vehicle model year (Transportation Data Book (1994) or MVMA (1995)), and most existing models of fleet emissions make simple assumptions about the impact of policies or changes in prices on the number of vehicles scrapped and their underlying characteristics, such as their expected remaining life.[4] However, to evaluate the costs or welfare implications of policies that encourage fleet turnover, it

[3] In Harrington, McConnell and Alberini (1995) we examine the influence of technical characteristics of emission measurement and repair on the efficiency properties of various emission fee/subsidy policies. Harrington and Walls (1996) use the model to examine the distributional properties of various command-and-control and economic incentive programs to reduce in-use emissions.

[4] The most notable model is EPA's MOBILE model which forecasts fleetwide emissions. At the present time, this model has no scrappage component, but EPA has designed an old car scrappage regulation that assumes cars that are scrapped early would have had a three year average remaining life (EPA, 1993). A model developed by EEA (1994) to be used with EPA's MOBILE model does have scrappage component. In this model, all vehicles within any given vintage are assumed to be identical, i.e., they have the same value, same emissions, etc.

is important to know the distribution of vehicle values within model years. Cars that are scrapped early are likely to be those whose value to their owners is the lowest. When designing scrap policies, it is important to know the characteristics of the owners of the lowest valued cars and their vehicles.

We model the decision to scrap a vehicle drawing from our earlier work (Alberini, Harrington and McConnell (1995)).[5] We assume that a vehicle owner balances expected future marginal costs and benefits in making the decision about how long to hold the vehicle. Specifically, when the vehicle is first bought, the owner chooses the optimal ownership time by balancing expected future marginal costs and benefits. However, this initial decision is subject to re-evaluation at the beginning of each time period, as additional information becomes available about the intrinsic quality of the vehicle and the owner's demand for driving services.

At the beginning of each period, simultaneously with fixing the optimal ownership time t^* in which the marginal benefits equal marginal costs (possibly a revision of earlier estimates), the owner also determines the path of maintenance and repairs to be undertaken from the present to the end of the planned ownership time. The value of the car, or the owner's minimum willingness to accept (WTA) value for it, is equal to the present value of the stream of net benefits associated with owning and driving the vehicle. At time t^*, the owner scraps the vehicle if its value (WTA) is less than the scrap value minus the cost of the repairs that would be necessary to keep the car in working condition (Parks, 1977). If WTA exceeds the scrap value net of the cost of repairs, the owner will sell the vehicle to another individual, who will drive it for some time. As a result, the decision to scrap, sell or keep the vehicle in any period depends on the vehicle's value relative to its scrap value and the cost of repair.

Following our model of vehicle ownership, determinants of the vehicle's value, which are a function of both the marginal costs and benefits of holding the vehicle, are central to the scrappage decision. We estimate vehicle value using a unique longitudinal data set collected over three years as part of the Delaware Vehicle Retirement Program. The owner's willingness to accept or WTA is taken as the best

[5] In that study, we extend models by Park (1980) and Gruenspecht (1985) to derive a model of the optimal time to hold a vehicle.

measure of an individual vehicle's value and is regressed on cost and demand characteristics using a number of different econometric specifications. The estimated coefficients in the WTA model are used as inputs into our simulation models with a scrappage program component.

2.1 Determinants of Willingness to Accept

The owner's minimum willingness to accept for a vehicle is the present value of the current and future stream of net benefits associated with owning and driving the vehicle, and is formally defined as:

$$\text{(1)} \qquad \text{WTA}=\int_0^{t^*}[B(t)-C(t)]dt$$

where B(t) represents benefits at time t, and C(t) measures costs at time t. Willingness to accept is, therefore, determined by the path of marginal costs and benefits of holding on to the vehicle, and by the time at which marginal costs and marginal benefits are equalized.

The benefits B(t) associated with the vehicle (the demand for driving services, which can be thought of as the value of the miles driven, perhaps adjusted by a quality of driving index) depend on individual and household characteristics, such as household income, age of the owner, size of the family, how many cars the household owns relative to the number of household members or licensed drivers, and the need to use the car for work-related purposes.

We express the costs C(t) associated with the vehicle (maintenance and repairs) as a function of the age of the vehicle, its condition, past repairs, the total miles on the vehicles, the miles driven in the most recent period, and perhaps other factors related to emissions, such as the vehicle's waiver status. Note that all of these variables are predetermined (they are the results of decisions and repair expenditures undertaken in the past, but not the object of current decisions) or are outside of the owner's control (such as the age of the vehicle).[6]

To summarize, WTA can be written as:

$$\text{(2)} \qquad WTA = f(B(t^*),C(t^*)) = f(X)$$

[6] This is to ensure that the independent variables in our model of WTA are not simultaneously determined with the dependent variable, willingness to accept.

where X is a vector of exogenous variables, including household income, age of the owner, number of cars owned, number of licensed drivers in household, age of the vehicle, odometer reading, miles driven in the previous year, condition of the vehicle, expenditure in repairs and maintenance in the previous year, and waiver status. We turn relationship (2) into an econometric model by including an error term that captures the distribution of WTA in the fleet.

2.2 The Data

We obtained owner-assessed values for relatively old vehicles in the course of interviews of vehicle owners conducted in association with the Delaware Vehicle Retirement Program (DVRP). The DVRP took place in 1992 and is one of the most recent examples of accelerated vehicle retirement programs, whereby owners of old – and presumably highly polluting – vehicles are offered a bounty to give up their vehicles, which are then disposed of in an environmentally sound way.[7] Removing these vehicles from the fleet and replacing them with newer and cleaner vehicles is argued to reduce in-use emissions.[8]

The DVRP targeted approximately 4200 owners of pre-1980 vehicles, who were offered $500 to give up their vehicles. The targeted owners received letters that spelled out the nature of the program, the bounty level and asked interested owners to call a toll-free number in order to make arrangements for scrappage. One-hundred twenty-five vehicles were purchased, and 121 of the owners of those vehicles were interviewed at the scrapyard. A total of 365 non-participants (owners of pre-1980 vehicles who were sent letters soliciting participation in the program, but had chosen not to participate) were surveyed over the telephone, whereas the 48 "waitlisted" owners (owners who indicated they wished to participate, but had replied to the DVRP letters only after the goal of 125 vehicles had already been attained) were not interviewed in this first round of surveys.

---

[7] The DVRP was initiated by the U.S. Generating Co. Total Petroleum ran an accelerated vehicle retirement program in the Denver Metro area in 1994 (Lodder and Livo, 1994). The Total Petroleum program was somewhat different in that it included a scrap-or-repair component. There have been other examples of scrap programs in Illinois EPA (1993) and the UNOCAL program in California (see Dickson et al, 1991, and Tatsutani, 1991).

[8] Alberini, Harrington and McConnell (forthcoming) show that the extent of the emissions reductions depends crucially on how much longer those old vehicles would have been kept in use in the absence of the scrappage program, on the miles driven every year, and on the age of the replacement vehicle.

In this first round of surveys, both participants and non-participants were asked similar questions. Specifically, we verified the information on make and model year developed by the organizers of the program, asked whether the car had been purchased new or used by the current owner, inquired about the odometer reading, the miles driven in the previous year, the current use of the vehicles for commuting and non-commuting work-related purposes and errands, the present condition of the car and the maintenance expenditures in the previous year as well as those planned for the next year. In addition, we asked how much longer the owner planned to keep the vehicle, and how he or she was planning to dispose of it at that time (by selling, trading or scrapping it).

One of the most important questions was about the car value: we elicited information about willingness to accept for the vehicle by asking the owner if he or she would have participated in the scrappage program if the program's offer had been $X. The dollar value, $X, started at $400 for participants, since their willingness to accept must have been lower than $500, at least at the time the DVRP letter was received, and was subsequently lowered in follow-up questions by $100 at a time until the respondent declined to participate. The bounty at which the respondent declined to participate pegs his or her WTA value. The dollar value, $X, started at $600 for non-participants, since their willingness to accept must have been greater than $500, at least at the time the DVRP letter was received, and was subsequently raised in follow-up questions by $100 at a time until the respondent agreed to participate. The level of the bounty at which the respondent agreed to participate pegs this respondent's WTA amount.[9] The survey ended with questions about the household's economic circumstances and demographics.

This first round of surveys determined that a number of owners did not have their pre-1980 vehicle at the time the DVRP letters were received, and that a number of non-participating owners had either scrapped or sold their vehicles between the time of the DVRP letters and the time of the interviews.

[9]Those respondents who stated that they would not have participated in the scrappage program if the offer had been $1000 were asked to provide a point value for their WTA. Several respondents indicated that they would not have participated at $1000, but failed to provide a point estimate of their WTA value. We developed special statistical models to accommodate for these responses.

These persons were given an abbreviated version of the survey questionnaire that omitted the willingness to accept questions as well as many other questions about the recent use and condition of the car. We did, however, ask when and how the vehicle was disposed of.

About a year later, we once again contacted over the telephone the non-participants who still had their pre-1980 vehicle at the time of the first round of surveys. Those owners who still owned the vehicle at the present time were given a survey questionnaire that was essentially identical to that in our first round of surveys.[10] In addition, we contacted and interviewed most (42) of the "waitlisted" owners and interviewed them over the telephone.[11]

Finally, another year later we re-contacted all of those non-participants and "waitlisted" owners who had reported owning the car at the time of the second round of surveys and repeated the standard version of our questionnaire.

2.3 Econometric Specifications

The three round of surveys enabled us to develop a longitudinal dataset that includes (i) participants, (ii) non-participants who still owned their pre-1980 vehicle at the time of the first round of surveys, and (iii) "waitlisted" owners who still owned their pre-1980 vehicle when first surveyed. Owners (ii) and (iii) provide, at regular intervals of one year, information on the most recent condition, use, value and planned ownership for their car.[12] Because one owner's WTA value and planned ownership time often change from one survey to the next, we use our longitudinal dataset to estimate a model that relates WTA to the most recent condition, use and repair information.

---

[10]Those owners who had scrapped or sold their vehicles between the previous and the present survey were given the abbreviated version of the questionnaire.

[11]One important difference between the questionnaire we administered to the "waitlisted" owners and the standard questionnaire is that we asked questions about the "waitlisted" owners' use and value of the car at the present time as well as at the time the DVRP letters were first received by these owners. Once again, those owners who had sold or scrapped their car between the time of the DVRP letters and the present were given an abbreviated version of the questionnaire.

[12] Since owners drop out of our dataset as soon as it is ascertained that they do not hold their vehicles any longer, we have a minimum of one and a maximum of three observations per owner in the dataset. Those owners who still had their cars at the time of the most recent round of surveys contribute three observations.

The longitudinal structure of our dataset allows us to formulate a number of alternative specifications for our model of WTA. A first, basic specification allows WTA to be determined by all of the factors suggested above, does not include the previous survey's WTA among the independent variables, and assumes that observations from the same owner are fully independent of each other. Formally, the model for willingness to accept is:

$$\log WTA_{it} = x_{it}\beta + \varepsilon_{it} \qquad (3)$$

where i indexes the individual (i=1, 2, ..., n), t indexes the round of surveys (t=1, ..., T_i, where T_i may be equal to one, two or three, depending on the fate of the respondent's vehicle), x is a vector of determinants of WTA (individual or household characteristics; vehicle characteristics), and ε_{it} is an error term. Cov($\varepsilon_{it}, \varepsilon_{js}$) is set to zero for t≠s and all i's and j's. We choose log WTA as our dependent variable because previous work with the data from the first-round surveys suggests that WTA is reasonably approximated by a log normal distribution (Alberini, Harrington and McConnell, 1995).

Our second specification postulates that WTA is serially correlated, and takes care of the serial correlation by including last year's WTA value among the independent variables. Formally, the model of willingness to accept is:

$$\log WTA_{it} = x_{it}\beta + \gamma \log WTA_{i,t-1} + \varepsilon_{it} \qquad (4)$$

with the errors independent of each other for all persons and rounds of surveys. The usable sample size for this specification is necessarily reduced by the fact that we can only use individuals who reported exact WTA values in at least two round of surveys.

The nature of some of the responses to the willingness to accept questions prevents us from using least squares when estimating our models of willingness to pay. We resort to maximum likelihood to accommodate those respondents who – in one or more rounds of surveys – declined to participate in the scrappage program at \$1000 but never reported their exact WTA value.[13] The log likelihood function is:

[13]The contribution to the likelihood is the probability that willingness to accept (a log normal variate centered around xβ) is greater than 1000.

$$\text{(5)} \qquad \log L = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \left[(1 - I_{it}) \cdot \log \varphi(\log WTA_{it}; x_{it}, \beta, \sigma) + I_{it} \cdot \log \Phi\left(\frac{x_{it}\beta}{\sigma} - \frac{\log 1000}{\sigma} \right) \right]$$

where $\phi(\bullet)$ and $\Phi(\bullet)$ denote the standard normal pdf and cdf, respectively; σ is the standard deviation of the error term, and I_{it} is an indicator that takes on a value of one for those respondent who would not have participated in the program for $1000, but do not report their exact WTA value, and zero for all others.[14]

A description of the variables used in our regression is provided in Table 1, and regression results for the first and second specifications are reported in Table 2.a and 2.b, respectively.

Table 1.
Description of variables from the DVRP surveys.[15]

Variable	description	mean	std devn	min	max	#valid
WTA	exact WTA value	1535.58	2791.27	100	20000	506
age2_ye	age of the car in years	17.13	2.95	13.5	32	848
miles	odometer miles	126,060	61,415.15	1000	430,467	457
pastyr	miles driven in the past year	4343.92	3504.57	1000	12,000	543
wvalue	Blue Book Value at time of first survey	1021.83	1046.61	100	9850	522
cond	dummy that takes on a value of 1 if vehicle is in fair/poor condition; 0 for excellent/good condition	0.55	0.50	0	1	861
spent	how much money was spent to keep the car running in the past year	217.91	137.18	100	600	537
spend	how much money is to be spent to keep the car running another year	187.52	143.66	100	600	537
unantic	unanticipated repairs	-58.28	179.61	-500	500	163
income	household income	36,663.75	19,985.50	10,000	75,000	571
owned	number of vehicles owned by the household	2.77	1.42	0	16	805
liscdriv	number of licensed drivers in the household	2.09	0.90	0	6	856
waiver	a dummy variable that takes on a value of 1 if the vehicle has been granted waiver status, 0 otherwise	0.34	0.47	0	1	861
age of owner	years	49.48	15.95	18	92	807

[14] The log likelihood function here reported, (5), refers to model (3) of willingness to pay. It is modified to reflect model (4) of willingness to pay.
[15] Other variables used in the WTA regressions: lmiles = log odometer miles; lpastyr = log miles driven in the past year; lwvalue = log Blue Book value; lspent = log(spent); lincome = log household income.

2.4 Econometric Results

Among car characteristics, we expect higher odometer mileage, older age and poor condition to decrease the value of the car. Waiver status may also decrease the value of the car, whereas the effect of past maintenance and repairs is uncertain a priori: high maintenance expenditures in the past may imply that this vehicle has been taken good care of, but may also signal a poor quality car.

We initially ran regressions that included both the vehicle characteristics (which drive costs) as well as individual/household characteristics (which we assume are the main determinants of the benefits of owning the vehicle). However, individual and household characteristics were never found to be significant in the models of WTA that included both individual/household characteristics and vehicle

Table 2.a.
WTA model: specification (1). Dependent variable: log WTA.
(T statistics in parentheses).

Indep. variable	(A)	(B)	(C)	(D)	(E)
intercept	12.6836 (9.057)	2.9667 (2.353)	5.4282 (2.586)	6.7518 (2.983)	7.4719 (25.311)
age2_ye	-0.0540 (-1.734)	0.0078 (0.299)	0.0461 (1.088)	0.0503 (1.199)	-0.0333 (-1.957)
lmiles	-0.3491 (-2.895)				
lpastyr	-0.1279 (-1.700)	-0.0363 (-0.665)	-0.0810 (-0.853)	-0.0594 (-0.627)	
cond	-0.8264 (-6.013)	-0.7683 (-7.926)	-0.8382 (-3.965)	-0.7531 (-3.488)	
lwvalue		0.5847 (5.506)	0.2615 (1.269)	0.2561 (1.259)	
lspent	0.0999 (0.987)	0.0915 (1.186)		-0.3147 (1.463)	
waiver	-0.1000 (-0.706)	0.0252 (0.216)	-0.1902 (-0.823)	-0.1865 (-0.819)	
unantic			-0.0015 (-3.238)	-0.0025 (-3.116)	
owned		0.0068 (0.199)			
stand devn of error	1.1729 (14.833)	1.0247 (17.313)	0.9937 (9.836)	0.9790 (9.860)	1.2048 (20.301)
sample size	344	404	121	121	632
log L	-216.33	-267.64	-80.76	-79.90	-458.02

characteristics,[16] so we report results for those regressions that only include determinants of costs among the regressors.

Table 2.a shows the results for the first specification of the econometric model. As shown in Column (A), as we expected, age tends to depress the value of the vehicle (the coefficient of age being negative and significant at the 10% level), but the effect of age is dominated by that of odometer miles (which tend to be correlated with age, and have a negative and highly significant coefficient) and condition of the car. The miles driven in the previous year also tends to correlate negatively with WTA (the coefficient being significant at the 10% level). Waiver status does not seem to affect the values of the car. The coefficient of past maintenance is positive, but not significant at the conventional levels.

In Column (B) we eliminate odometer reading and include blue book value at the time of the first survey.[17] Blue book is, in fact, one of the strongest predictors of WTA, the other being its condition. This suggests that the owner-assessed value of the vehicle tends to follow the market average for vehicles of that model year, the difference relative to this average being explained by the condition of the vehicle "for its age." In addition, in the span of time covered by our surveys (about two years) the present condition of the vehicle is sufficient to explain the decline in value relative to the initial-survey blue book value: the miles recently driven and age have no explanatory power of their own, suggesting that miles and age are correlated with condition.

In columns (C) and (D) we included UNANTIC, the unplanned component of maintenance. UNANTIC is computed as the difference between the planned future expenditure reported in the earlier survey and the actual expenditure in the year immediately preceding the current round of surveys. A large, positive value of UNANTIC means that the car has been undermaintained relative to what earlier planned, and is associated with a lower vehicle value. Essentially, moving from zero unanticipated

[16] These results are consistent with those reported by Morey (1996), who analyzes willingness to accept data from the Total Petroleum scrappage program.

[17] We know from previous investigations (Alberini, Harrington and McConnell, 1995) that blue book value is predicted from age and odometer reading. We include current age and miles driven most recently to proxy for what the blue book value would be at the time of the second and third round surveys.

expenditure to a level of $100 implies a decline in value of about 14%. Interestingly, the sign of actual past maintenance (LPASTYR) changes to negative when this variable is included together with the unanticipated component of maintenance, but is not fully significant at the conventional levels.

Column (E) of table 2.a isolates the effect of vintage alone, which is important for the simulations discussed in the remainder of the paper. Age *is* a significant predictor of WTA, its coefficient being negative and significant at exactly the 5% level.

Finally, Table 2.b shows that today's value is strongly correlated with the vehicle value reported by the owner in the immediately preceding round of surveys, all other variables (miles driven between the two surveys, present condition of the car, etc.) offering no additional explanatory power.

Table 2.b.
WTA model: specification (2). Dependent variable: log WTA.
T-statistics in parentheses.

Variable	(A)	(B)
intercept	1.3608 (0.649)	1.6233 (1.778)
log WTA of previous survey	0.6489 (4.539)	0.7499 (5.968)
age2_ye	0.0169 (0.244)	
lpastyr	-0.0149 (-0.128)	
lspent	0.1777 (1.136)	
waiver	-0.1880 (-0.666)	
cond	-0.3169 (-1.289)	
stand devn of error	0.9167 (8.626)	0.9468 (8.418)
sample size	87	88
log L	-52.37	-48.45

3. A SIMULATION MODEL OF VEHICLE INSPECTION AND REPAIR

The results of the WTA regressions are used along with the model of scrappage to incorporate old car scrappage into a simulation model of fleet emissions that includes stochastic and behavioral elements of emissions measurement and repair. The model allows vehicle owners to scrap vehicles if the cost of repair exceeds the owner's reservation price for the vehicle (WTA), less the scrap value. The scrap value is the bounty offered in an old car scrap program operating either in isolation or in conjunction with an I/M program or vehicle emissions fee. In the absence of an old car scrap program, the scrap value is simply the value of scrap metal and old car parts, and the simulation model effectively analyzes I/M and vehicle emissions fee policies alone.

The simulation model creates a "virtual" fleet consisting of approximately 1000 vehicles with an age distribution similar to the age distribution of vehicles observed in use in California in 1991 (EEA 1994). Each of these vehicles has been assigned an initial "true" rate of hydrocarbons (HC), carbon monoxide (CO), and nitrogen oxides (NOx) emissions, expressed in grams per mile (g/mi.). The simulated emissions of HC and CO are randomly drawn from a distribution of on-road emissions estimated using data from 90,000 vehicles in California that were subject to remote sensing of emissions in 1991.[18] The distribution is assumed to be bivariate log-normal, with means and variance-covariance matrices equal to the model-year-specific sample means and sample variance-covariance matrices estimated in the remote sensing study.[19] The simulated NOx emissions were drawn from vintage-specific distributions of I/M240 emission test results taken at EPA's emission test research facility in Hammond, Indiana. NOx emissions are assumed to be independent of HC and CO.

[18]Remote sensing is a technology combining roadside monitors that send infrared beams from one side of the road to a detector on the other side, measuring a vehicle's emissions, with a video camera that obtains a photograph or electronic identification of the license plate. See Stedman, et al. (1992), for a discussion of remote sensing technology and the California data used in this study.

[19] The high correlation of CO and HC emissions necessitates a bivariate distribution, rather than independent distributions for each pollutant.

3.1 Simulation Of The Regulatory Program: I&M

The I&M program is characterized by a set of model year-specific cutpoints, one for each of the three pollutants, HC, CO and NOx, which determine whether the vehicle passes the test; these cutpoints also vary with whether the vehicle is a car or light-duty truck. We use as a set of "base" cutpoints the cutpoints in actual use in California in 1992, as reported by Klausmeier et al. (1994). Each simulated "vehicle" proceeds through the simulation in the following steps:

1. Initial vehicle emission measurement. Current emission tests measure true emissions with some measurement error. We base the measurement accuracy on a comparison of emission measurements obtained from California's current BAR-90 test to measurements made with the Federal Test Procedure (FTP) on the same vehicles (California I/M Review Committee, 1993). (The FTP should measure true emissions.) These regression results are shown in Appendix Table A.2. An emission measurement is drawn for each vehicle based on the error distribution in this regression. We simulate a somewhat, but not excessively, imperfect test accuracy by multiplying the standard error by a parameter κ, equal to 0.5.[20]

2. Owner response: Any vehicle that has measured emissions exceeding any cutpoint is subjected to repair, or, at the owner's option, retirement. The simulated cost of each repair is a random draw from a log-normal distribution estimated from the reported repair costs in an 1100-car study by the California I/M Review Committee (1993).[21] We also use the data from this study to estimate post-repair emissions as a function of pre-repair emissions and repair costs. Owner reservation prices, used to determine whether to scrap the vehicle, are drawn from vintage-specific log-normal value distributions with parameters estimates as shown in Table 2.A, column (E), above. The model terminates here if the vehicle is scrapped. If the vehicle is repaired, post-repair emissions are determined by applying to the true

[20] For comparison, κ=0 represents BAR-90 test accuracy and κ=1 represents perfect accuracy – i.e., FTP results.

[21] Average repair costs from this study were approximately $89 per vehicle and are held down by a model year-dependent cost cap designed to prevent exceptionally large impacts on particular motorists.

emissions the coefficients in Table A.4 in the Appendix, plus a random error distributed as the errors in that table.

3. Retest. A second draw is made from the post-repairs emission test distribution in Table A.4.

4. Second repair. If any cutpoint is still exceeded, the model proceeds to the second repair. This repair is allowed to be more extensive than the first repair. The simulated costs and emission reductions are parameterized using data from a vehicle repair study conducted by the Sun Oil Company.[22] See Table A.5 in the Appendix. Again, the owner has the option of scrapping the vehicle instead of repairing it.

5. Stop. If the vehicle remains unable to pass the test after the second repair, it is allowed to be operated even though it is in violation.

3.2 Simulation Of The Economic Incentive Program: Emissions Fee.

The economic incentive program uses many of the same elements as the regulatory program. The main difference is in the importance of predicting the reduction in emissions, for the decision to repair the vehicle is based on this prediction. The following steps are developed:

1. Initial vehicle emission measurement and calculation of initial fee:

$$Fee_0 = \max\left\{0, \sum t_i(\hat{E}_i - Baseline_i)\right\}$$

where t_i is the fee rate for each pollutant, $\hat{E}_i$ the measured emission rate (as in the command-and-control policy above), and *Baseline* the level of "free" (tax-exempt) emissions granted each vehicle, if there are any.

2. Prediction of emission reductions and estimation of post-repair fee:

$$EstFee_1 = \max\left\{0, \sum t_i(\widetilde{E}_i - Baseline_i)\right\}$$

[22] We assume that motorists would first implement relatively inexpensive repairs, much like the ones observed in the California repairs dataset. If the vehicle still fails the emissions test, the owner would switch to more expensive repairs, like those reported in the Sun Co. study dataset.

where $\tilde{E}$ is predicted emissions using the repair cost and effectiveness from the California I/M Review Committee study. The predicted repair is based on the regression whose results are reported in Table A.4 of the appendix -- i.e., post-repair emissions are a function of pre-repair emissions, estimated repair cost, and vehicle model year. We assume that owners have a reasonable, but not perfect, ability to predict the emissions reduction attained through the repairs.[23]

3. Compare Fee_0, Repair cost $+EstFee_1$ and Owner's WTA, net of scrap value. If Fee_0 is smallest, do nothing. If Repair cost $+EstFee_1$ is smallest, repair. Otherwise, scrap. "Scrap value" is replaced by the bounty offered to owners of older cars as an inducement to vehicle retirement when we consider emissions fees combined with a scrappage program.

4. Repeat 2 and 3, using the Sun Oil Co. repair cost and effectiveness (see Table A.5 in the appendix).

4. RESULTS

We use the simulation model to examine a number of alternative policies affecting a 1,000-car representative California fleet. We examine scrappage rates, repair rates, net benefits and cost effectiveness under alternative policies. We focus on three policies:

- *Old car scrap.* A stand alone scrap program in which cars are purchased at a specified offer price. The offer price is allowed to vary from \$100 to \$1,000;
- *Regulatory Program (with and without scrap program).* A regulatory program which represents the current generation of state I/M.

[23] We simulate improvement in repair-effectiveness prediction with a parameter λ defined in the unit interval, with λ=0 representing the predictive ability shown in Table A.4 and λ=1 representing perfect predictability. For the simulation model of this paper, λ is set to 0.5. We allow both λ and κ to vary in a related paper (Harrington, McConnell, and Alberini, 1995) in which we discuss changes in technical parameters of the model, the precision of the emissions measurement (variations in κ) and the ability to predict repair effectiveness (variations in λ), and use the simulation model to examine the effect on emission reductions and costs of changes in the fee rates, cutpoints and other policy parameters. Harrington and Walls (1996) use a similar model to examine the distributional implications of various in-use vehicle emission policies.

- *Emissions fee policy (with or without scrap program).* Emissions fees on vehicles have been suggested by many economists and policy makers (White (1982), Kessler and Schroeer (1993) and Harrington, McConnell and Alberini (1995)), but have yet to be implemented. They are, however, currently under consideration in California.

The emission fee we use in this analysis would require that emissions of all cars be tested, and that owners would be required to pay a fee based on the gram per mile emissions test results. The fee is in cents per gram/mile and for this analysis has been set equal to the marginal damages as measured by Small and Kazimi (1994): 0.3 cents per gram mile for HC and 1 cent per gram per mile for NOx. A scrap program is added to the fee program, giving owners an alternative to repairing the car or paying the fee.

Two types of fees are considered. The first type has no exempt emissions (no baseline: owners must pay the fee on all emissions), whereas the second type (baseline 1) owners must pay the fee only on emissions greater than some allowed level, which we set equal to the average emissions level of the fleet.

Figures 1a and 1b show the results of including a scrap program with an emissions fee and an I&M program. First, with no structured old car scrap program (bounty equal to $0), about 2% of the fleet is scrapped each period due to both average fleet turnover and the presence of the I/M program and its required repairs. With the emissions fee policy (no baseline) set at the level of marginal damages as described above, with no scrap offer, about 7% of the fleet is scrapped. As a scrap program is introduced, and the offer increases up to $1,000 per car, the number of cars repaired declines and the number scrapped increases. Increases in the scrap offer cause drivers to elect to scrap instead of repair. Those cars that are scrapped will be the ones that have the highest repair cost relative to the driver's valuation of the car.

Figure 2 addresses the broader issue of the net benefits under different policies as the scrap offer varies from 0 to $1,000. Benefits are calculated as the emissions reductions under each policy times a constant marginal damage function. We use marginal damages estimates from Small and Kazimi (1994)

Figure 1a
Cars Scrapped and Repaired Regulatory Program

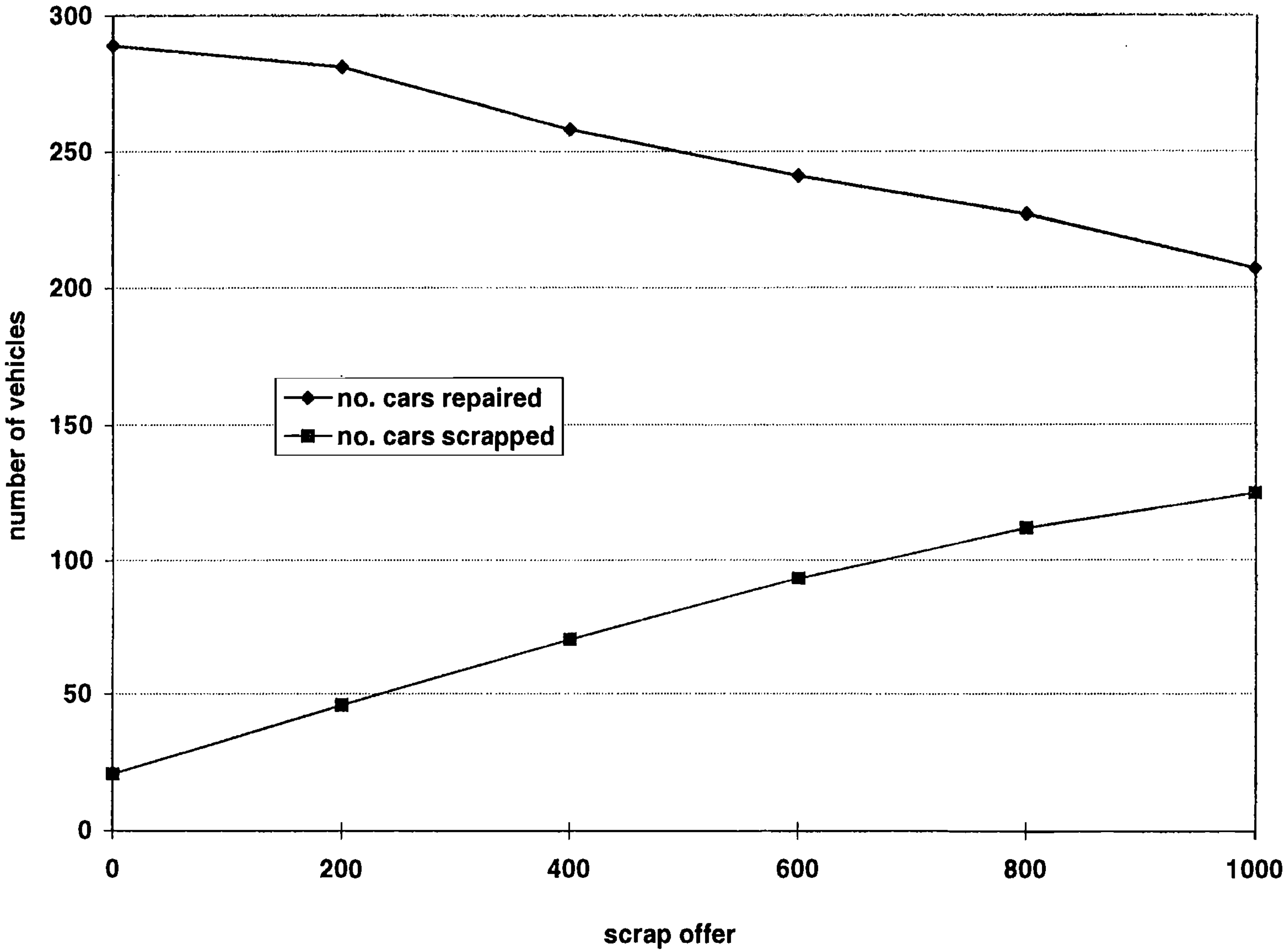


Figure 1b

Cars Scrapped and Repaired Under Emissions Fee Program, Zero Baseline

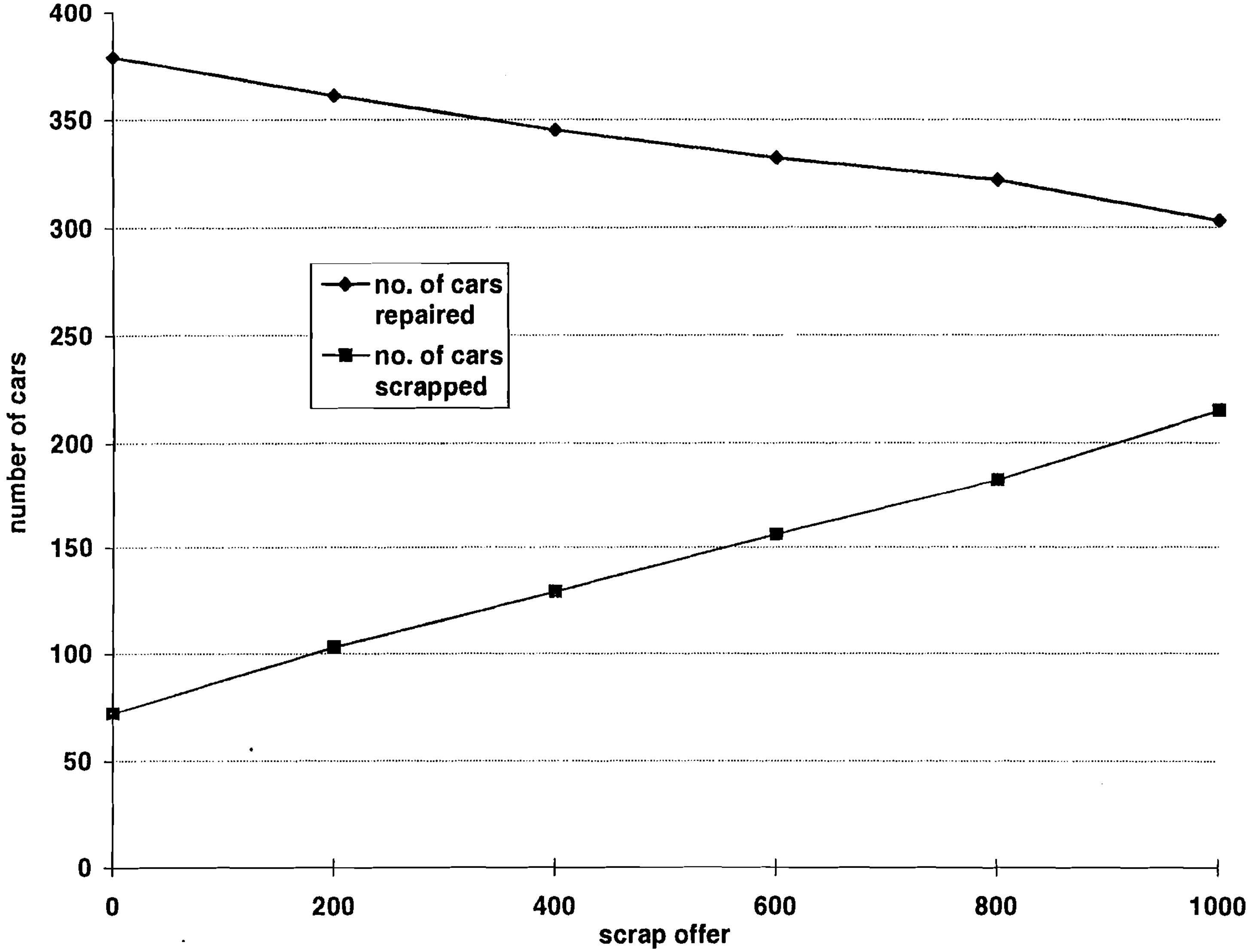
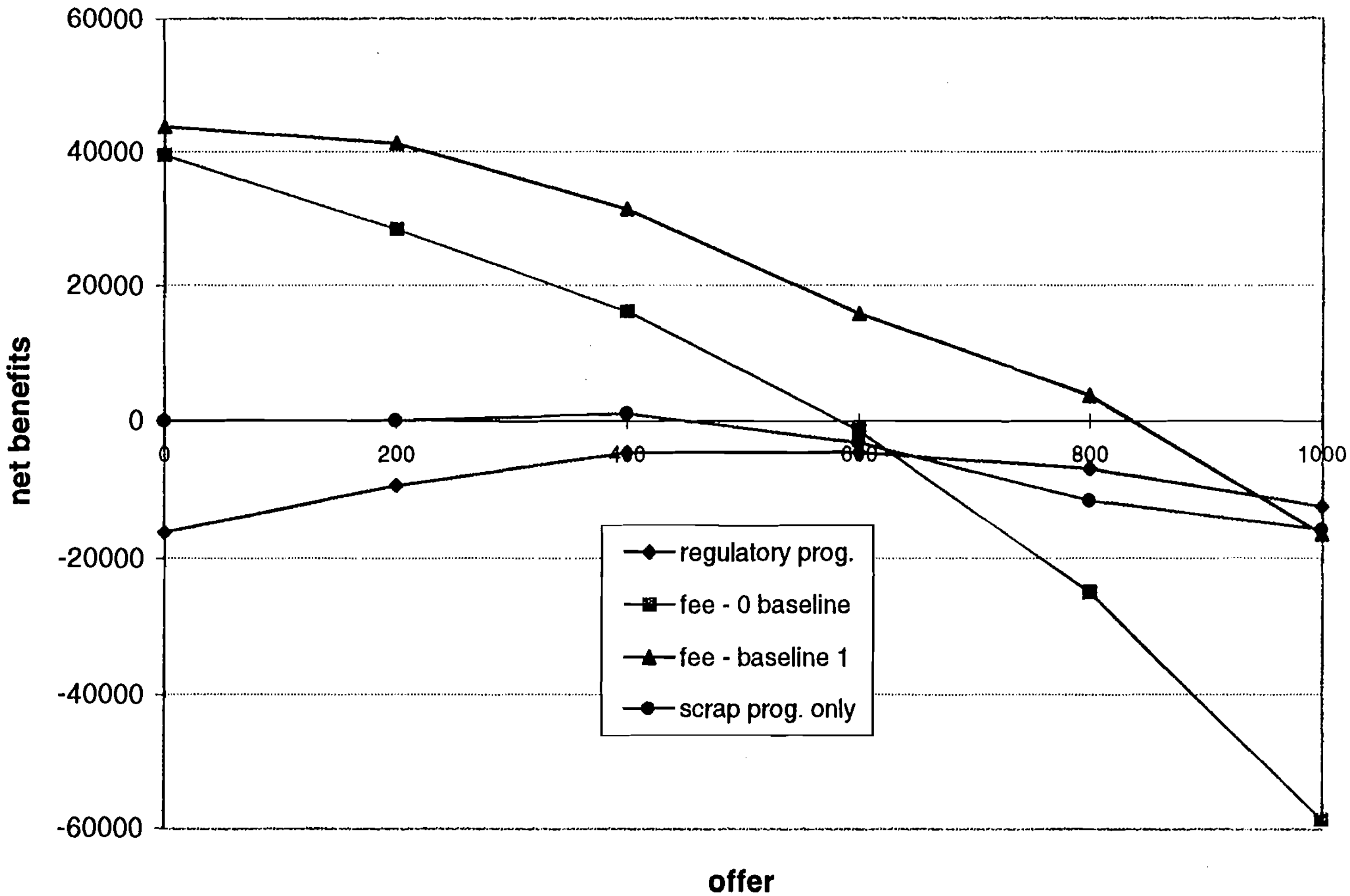


Figure 2

Net benefits for different policies - with varying scrap offers



of \$3,000 per ton of hydrocarbons and \$10,000 per ton of nitrogen oxides.[24] The costs of each program are calculated as the repair cost, plus the cost of scrapping cars early (the value of the cars that get scrapped). We do not count the fee payment as part of the social cost in the case of the emissions fees, assuming that taxes will be reduced by an equal amount elsewhere. Similarly, with the scrap program bounty, we assume that taxes must be raised elsewhere to raise the money to make the bounty offers, so there is simply a redistribution of funds.[25] Finally, we do not include the testing costs under the regulatory or fee policy, because we assume these costs would be the same under either policy.

As we would expect, Figure 2 shows that with no separate scrap policy (or \$0 offer), the regulatory and simple scrap programs have net benefits well below the fee policies. The fee is set at the optimal level, at the level of marginal damages, inducing only the most cost effective repairs. The I/M program, on the other hand, requires all repairs to be made regardless of the value of the emissions reduced relative to costs. When a scrap program is introduced, as shown in Figure 2, net benefits for the regulatory I/M program increase, while a scrap program decreases the net benefits of either emissions fee program. Under the I/M regulatory policy, an scrap offer between \$400 and \$600 maximizes net benefits. However, net benefits remain negative at all bounty levels, implying that the social costs are not worth the emissions reduction benefits. At offers below the optimal level, the value of the cars scrapped must be less than the value of the emissions reduced from scrapping them. Under the fee policies, scrap offers must cause cars to be scrapped whose value in use exceeds the value of the emissions reductions from scrapping. Hence, net benefits decline as the scrap offer rises.

The scrap program alone generates modest benefits up to a bounty of a little over \$400. Higher bounty offers result in the scrappage of vehicles that have higher in-use value than the value of eliminated

---

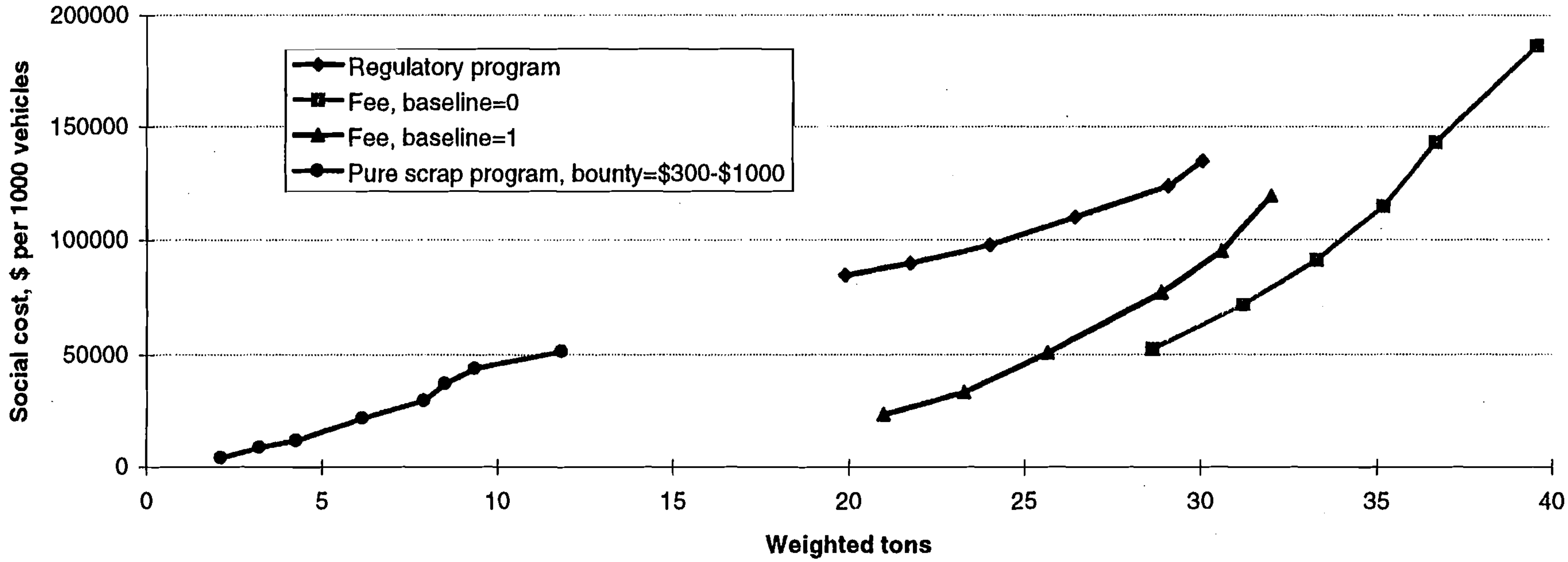
[24] The damages from vehicle emissions include the increased morbidity from ozone formed by HC and NOx emissions and the increased mortality caused by some types of particulates including some NOx particulates. Small and Kazimi (1994) find that the mortality effects from nitrogen oxides are about 3 times higher than the damages from hydrocarbons. This estimate, especially for NOx emissions, is high relative to other estimates of air pollution damages (see Krupnick and Portney, 1991).

[25] The costs and subsidies are, however, very real to drivers (see Alberini, Harrington and McConnell, 1996).

emissions, making the losses larger than the emissions reduction benefits. Finally, emissions fee policies prove superior to the other alternatives considered, especially for the variant without exempt emissions.

Figure 3 shows the total social cost functions for reducing emissions under the alternative policies, for different scrappage offer prices. Each point on the line for each policy represents a different offer price, from 0 to $1,000. Emissions reduced are given in terms of weighted tons reduced, where tons are weighted according to the marginal damages from HC and NOx (NOx emissions are weighted just over 3 times higher than HC). The stand alone scrappage program is relatively low cost, but provides relatively few emissions reductions, a finding that confirms the evidence from other analyses (Alberini et al., 1994). The regulatory program, on the other hand, attains a much higher level of emissions reductions, albeit at a higher cost. Of the two emissions fee program, the one with exempt emissions has emissions reduction potential comparable to that of the I/M program, but is much less costly. Adding a scrap program with successively higher scrap offers raises cost more quickly with the fee policies than with the I&M policy. This is because the low-valued high emitting cars have already been scrapped under an emissions fee policy, so scrapping additional cars will bring in higher valued cars.

Figure 3
Social costs and tons reduced
Scrap offer varies from 0 to $1000



5. CONCLUSIONS

We have examined the factors affecting owners valuations of their old vehicles using a unique longitudinal dataset. Willingness to accept is well predicted by the owner's valuation in last period, and on the mileage and condition of the car. Our estimated model of vehicle value is used as an input into a simulation model of a 1,000-car fleet representative of California's fleet. Other inputs into the simulation models are the estimated distributions of emissions in the fleet, and two equations that link emissions reductions to the cost of repairs.

The simulation model is used to examine the role of scrap policies alone and combined with other policies for reducing emissions, such as current state I/M program and proposed emissions fees, and the welfare implications of combining such programs. The model assumes that of all possible options (repairing the car, scrapping the vehicle or turning it in to an old car scrap program, paying the emissions fee without repairing the vehicle) the owner will choose the one with the least cost.

We find that old car scrap program may increase net welfare under a regulatory program like I/M in practice today, but that a stand alone scrap program is unlikely to provide very much in the way of emission reductions.

REFERENCES

Alberini, Anna, David Edelstein, Winston Harrington and Virginia McConnell, 1994, "Analysis of Emission Reductions and Cost of the Delaware Vehicle Retirement Program," Report prepared for the President's Commission on Environmental Quality, Resources for the Future Discussion Paper, Washington, DC (August).

Alberini, Anna, Winston Harrington and Virginia McConnell. 1995. "Determinants of Participation in Accelerated Vehicle Retirement Programs", *The Rand Journal of Economics,* Vol. 26, No. 1, Spring.

Alberini, Anna, Winston Harrington and Virginia McConnell. Forthcoming. "Estimating an Emissions Supply Function from Accelerated Vehicle Retirement Programs", *The Review of Economics and Statistics.*

Alberini, Anna, Winston Harrington and Virginia McConnell. 1996. "Fleet Turnover and Old Car Scrap Policies", paper presented at the AEA annual meetings, San Francisco, January.

Crandall, Robert W., Howard K. Gruenspecht, Theodore E. Keller and Lester B. Lave. 1986. *Regulating the Automobile* (Washington, DC: The Brookings Institution).

Dickson, Edmund L., R. C. Henning and W. R. Oliver. 1991. "Evaluation of Vehicle Emissions from the UNOCAL SCRAP Program," Radian Corporation, Sacramento, California, (June).

Gruenspecht, Howard K 1982. "Differentiated Regulation: A Theory with Applications to Automobile Emissions Control," Ph.D. dissertation, Yale University.

Hahn, Robert W. 1993, "An Economic Analysis of Scrappage," American Enterprise Institute Discussion Paper.

Harrington, Winston and Margaret Walls. 1996. "Controlling Vehicle Emissions: The Efficiency and Equity of Taxes versus Standards," paper presented at the AEA annual meetings, San Francisco, January.

Illinois Environmental Protection Agency. 1993. "Pilot Project for Vehicle Scrapping in Illinois", Springfield, Illinois, (May).

Kessler, Jon and William Schroeer. 1993. "Meeting Mobility and Air Quality Goals: Strategies that Work," U.S. Environmental Protection Agency, Office of Policy Analysis, Final Draft, October.

Klausmeier, Rob, de laTorre Klausmeier Consulting, Inc. 1995. "Evaluation of the California Pilot Inspection/Maintenance (I/M) Program," prepared for California Bureau of Automotive Repair, with Radian Corporation, Austin, Texas, March 31.

Krupnick, Alan J. and Paul Portney. 1991. "Controlling Urban Air Pollution: A Benefit Cost Assessment," *Science*, vol. 252 (April). pp. 522-28.

Lodder, Tymon and Kim Bruce Livo. 1994. "Review and Analysis of the TOTAL Clean Cars Program," Regional Air Quality Council and the Colorado Department of Public Health and Environment, Denver, Colorado, (December).

Manski, C.F. and Ephraim Goldin, 1983, "An Econometric Analysis of Automobile Scrappage," *Transportation Science,* 17 (4).

Morey, Edward. 1996. "Combining Responses to Actual and Hypothetical Offers to Estimate WTA for Highly-Polluting Clunkers: A Hypothetical-Bias Ordered-Probit Model of WTA?" paper presented at the AEA annual meetings, San Francisco, January.

National Research Council. 1991. *Rethinking the Ozone Problem in Urban and Regional Air Pollution,* (Washington, DC: Committee on Tropospheric Ozone Formation and Measurement, National Academy Press).

Parks, Richard W., 1977. "Determinants of Scrapping Rates for Postwar Vintage Automobiles," *Econometrica,* vol. 45, no 5 (p. 1099-1115).

Small, Kenneth and Camilla Kazimi. 1994. "On the Costs of Air Pollution From Motor Vehicles," University of California-Irvine Economic Paper No. 94-95-3.

South Coast Air Quality Management District. 1993. "Rule 1610: Old Vehicle Scrapping," (January 8).

Tatsutani, Marika. 1991. "UNOCAL Corporation's SCRAP: An Experiment in Corporate Environmental Initiative," Energy and Resources Group, University of California at Berkeley, (June).

U.S. Department of Energy. 1995. "Transportation Energy Data Book: Edition 15," Stacy C. Davis, Oak Ridge National Laboratory, ORNL-6856, (May).

U.S. Environmental Protection Agency. 1993. "Guidance for the Implementation of Accelerated Retirement of Vehicles Programs," Office of Mobile Sources, Ann Arbor, Michigan, (February).

Walls, Margaret and Jean Hanson, 1995. "Measuring the Incidence of an Environmental Tax Shift: The Case of Motor Vehicle Emissions Taxes," Draft Working Paper, Resources for the Future, Washington, DC.

White, Lawrence J. 1982. *The Regulation of Air Pollutant Emissions from Motor Vehicles.* (Washington, DC: The American Enterprise Institute).

APPENDIX
EMPIRICAL STUDIES USED IN THE SIMULATIONS

The simulation model relies on data taken from four empirical studies: the California I/M Review Committee "undercover car" study, and the Sun Co. scrap-repair study (Sun Co., 1994), the , and a study of on-road vehicle emissions using remote sensing (Stedman et al. 1992). The data are used to estimate statistical models of emission measurement and vehicle repair that generate the parameters of the probability distributions used in the simulation.

A1. Empirical studies of I/M test results

Until quite recently the standard methods for determining vehicle emissions in I/M test lanes have involved the sampling of tailpipe emission concentrations under"no-load" conditions. An example is the BAR-90 test used by the Smog Check program in California. These no-load tests have been criticized their inaccuracy, especially for late-model cars. "Accuracy" in this context refers not to the correspondence between the test results and the true emissions from the tailpipe during the test, a correspondence that depends on how well testing equipment is calibrated and whether the test protocol is observed. By this standard, no doubt, a well-done idle test can be quite precise. "Accuracy" refers do a more demanding standard: How well does the test result characterize the actual emissions of the vehicle under typical driving conditions? That is, how representative are the results?

To evaluate the performance of the BAR-90 test we use a data set collected by the California I/M Review Committee in 1992 as the central part of an assessment of the performance of California's "Smog Check" program. This program involved the recruitment of a large number of vehicles in use which were given an initial FTP test.[26] Those vehicles that failed the FTP were sent out to a sample of Smog Check stations in Southern California as if they were ordinary cars out to get their required Smog Check certificates. These "undercover" cars were given emission tests by the (presumably) unsuspecting Smog Check stations and if failing, were repaired and retested. The cars were then given a post-repair FTP test. Of approximately 1100 vehicles originally included in the program, we work with a data set of 681 vehicles for which repairs were attempted and a second FTP completed.

The emission test used by Smog Check stations during this period was the BAR-90 test, a two-speed test that includes, in addition to the idle test, a test of emissions at an engine speed of 2500 RPM.

Comparison of idle test and FTP test. The performance of the idle test was evaluated by regressing the FTP test results:

$$FTP = \alpha_0 + \alpha_1 IdleHC + \alpha_2 IdleCO + \varepsilon. \qquad (1)$$

where *FTP* refers to the FTP test emissions for HC, CO and NOx, in grams per mile, *IdleHC* the HC idle test results in ppm and *IdleCO* the CO idle test results in percent CO.

[26] FTP stands for "Federal Test Procedure" and is used in the U.S. for certifying emissions of new vehicles. In the major part of this test, the vehicle is placed on a dynamometer and tailpipe emissions are collected while the vehicle is put through a one-hour driving cycle designed to simulate urban driving patters. Although the representativeness of this driving cycle has been questioned, the FTP is a useful standard for measuring the performance of other emission tests and is taken in this report to represent the "truth."

Results, with standard errors in parentheses, are given in Table A1 below. As shown, the idle test results are completely ineffective at explaining FTP results for NOx (not surprising since NOx is not measured) and explains about a third of the variation in emissions of HC and CO. The most important parameters are the standard errors of the regressions, for these are used to generate the random normal deviates that serve as emissions measurement errors.

The regressions in Table A1 may shed an interesting light on a pair of other issues. First, the idle test is supposed to do a better job at predicting emissions of older vehicles (before widespread use of electronically monitored fuel injection), but that expectation is not borne out in the data. A Chow test failed to find any significant difference between pre- and post-1984 vehicles; in fact, if anything the performance of the idle test on newer vehicles was slightly better.

Another concern about the performance of the idle test is not borne out in these data. The performance of the idle test in an I/M context would be even worse than the poor fits below would suggest, because, if, as has been suggested, mechanics can fix the car to pass the idle test without actually reducing in-use emissions (Lawson 1993). That may be a compelling argument from an engineering perspective, but the effect does not show up here. If it were true, one would expect to see a difference in coefficients in the regressions of FTP on the idle test results using after-repair data compared to the regression coefficients using before-repair data. In fact, no significant difference is observed.

Table A1
Ability of the Idle Test to Predict FTP Emissions

	FTP HC	FTP CO	FTP NOx
Constant	1.305 (0.314)	25.36 (1.90)	2.09 (0.074)
Idle HC	0.00890 (0.00054)	0.0070 (0.0032)	0.37e-3 (0.12e-3)
Idle CO	0.279 (0.090)	9.93 (0.54)	-0.042 (0.021)
R-square	0.34	0.37	0.01
Standard error	5.86	35.52	1.38
N	669	669	669

Source: California I/M Review Committee, 1993. Standard errors in parentheses.

A2. Empirical studies of emission repairs

Recent results from the U.S. suggest that vehicle emission repair is not nearly as effective as had hitherto been assumed. In 1992 the EPA, in assessing the cost-effectiveness of the proposed Enhanced I/M regulation (USEPA 1993), assumed emission repairs to cost an average of $120 each and to have an effectiveness in reducing emissions as shown in Table A2.

Table A2 Repair Effectiveness Assumed by EPA, 1992			
Pollutant	**Emissions Before Repair (g/mi.)**	**Emissions After Repair (g/mi.)**	**Reduction (Percent)**
HC	0.74	0.41	45%
HC	1.92	0.61	68%
HC	3.9	1.08	72%
HC	14.3	1.38	90%
CO	9.27	4.99	46%
CO	28.0	9.47	66%
CO	90.0	12.1	87%
CO	190.7	20.6	89%

However, three recent studies indicate that vehicle emission repairs are much less successful and more expensive than assumed by the EPA. One of those studies was the 681-car data set described in the preceding section (California I/M Review Committee 1993). The average cost was low, less than $90 per vehicle on average, probably a result of waiver limits that restricted the amount spent on each vehicle. The average emission reduction was also low, 25 percent for HC, 21 percent for CO, and 8 percent for NOx. Emission reductions were also quite variable, with nearly half the cars showing higher emissions after repair than before, and emission reductions were not at all related to cost. The poor results may also be attributable in part to the way the program is administered by the BAR in California, where, according to some mechanics, it is easier to get in trouble with the state by presenting the motorist with a large repair bill than it is by failing to produce a genuine emission reduction.[27]

In two other recent repair studies the repairs were more effective, but far more expensive than assumed by the EPA. Both studies were conducted by oil companies searching for mobile source reductions to use as emission offsets for stationary source emissions. In 1993 Sun Oil Company (Sun Oil Co. 1994) used remote sensing to identify gross-emitting vehicles owned by their employees as they left company parking lots in Philadelphia. Sun offered to repair these vehicles at its expense, and the company spent up to $450 on each vehicle. As shown in Table 3, after average expenditure of $338, emission reductions for HC,CO and NOx were 68 percent, 75 percent and 9 percent, respectively. In terms of emission reductions, this is quite an improvement over the California results, but still a far cry from the EPA assumptions. Moreover, fully 40 percent of the vehicles were not brought into compliance even after expenditure of $450. It should also be pointed out that the vehicles were repaired by mechanics given special training. A second study was done by Total Petroleum in Denver. This was a combined scrap-repair study, in which gross-emitting vehicles were identified by several different means. The vehicles were split into "repair" and "scrap" portions, and the latter Total offered to repair at its own expense. As shown, emissions of HC and CO were reduced by about a third after average expenditure of nearly $400.

[27]The Bureau of Automotive Repair was originally established as a consumer protection agency, with the mission of fighting fraud in the auto repair industry.

Table A3 Repair Effectiveness in Non-EPA Empirical Studies				
	N	Average Cost	E_0 (g/mi.)	E_1 (g/mi.)
California I/M Review Committee (1993)	681*	$89.55		
HC			3.11	2.32
CO			37.21	29.4
NOx			1.67	1.53
Sun Company (1994)	155	$338.55		
HC			4.83	1.55
CO			69.18	17.00
NOx			2.90	2.64
Total Petroleum (1994)	103	$390.21		
HC			3.66	2.48
CO			45.64	33.38
NOx			not given	not given

Predictability of repair effectiveness. In a C&C I/M program it would be very useful to be able to predict by how much a given repair will reduce vehicle emission rates. If mechanics had this information *ex ante*, they could use it to select the combination of repairs that would bring the vehicle into compliance with the emission standards at least cost. Repair predictability would be even more useful in an emission fee program, since the mechanic can decide on the package of repairs that maximizes the expected net benefit of repair to the motorist -- that is, the expected reduction in emission fees paid less the cost of repair. This choice can include the possibility of making no repairs at all if it is less costly to pay the fee.

What would be most useful would be a repair model that gives the emission reductions that would follow from making certain kinds of repairs, or, equivalently, a model capable of predicting the excess emissions from a vehicle with a given set of conditions or malfunctions. Such a model would require data on the particular *types* of repairs done, and both the cost of repair and its effectiveness would depend on what is broken and what is done to fix it. Unfortunately the data are not available to estimate the effectiveness of specific repairs. The data available here include only the pre-repair emissions and the cost of a repair. Consequently, the regression model is not a causal model of repair effectiveness; the results merely express associations between post-repair emissions and a few pieces of information known pre-repair. If anything, the results may underestimate the mechanic's ability to predict the emission reductions from repair. When looking at a specific vehicle, the mechanic will have more information and will presumably be able to predict repair effectiveness at least as well as the models in the tables above.

We use data from the California I/M Review Committee and from the Sun Oil Co. study to estimate models of repair effectiveness. The basic model is

$$E_j^1 = \beta_0 + \sum \beta_i E_i^0 + \gamma_1 Age + \gamma_2 Cost + \delta \qquad (2)$$

where

$i, j = HC, CO, NOx$ are the pollutants of interest;

E^0, E^1 refer to emissions before and after repair, respectively

Age is the vehicle age in years,

Cost is the reported repair cost.

δ is the disturbance term, $\delta \in N(0, \sigma_r)$

As shown in Table A5, the results for California show that by far the most significant predictor of post-repair emissions of a pollutant are the pre-repair emissions of that same pollutant and indicates the marginal effectiveness of repair at removing pollutants from vehicle emissions. The coefficient of 0.35 for HC, for example, means that repair in the California program removed 65 percent of the incremental HC emissions. The effects of other pollutants are not consistently related to post-repair emissions. *Cost* is significant for CO and NOx and has the correct sign for all three pollutants, but in all cases the numerical magnitudes are so small that it has no practical importance. (For example, expenditure of $100 reduces expected HC emissions by 0.1 grams per mile.) *Age* is significant in the CO and HC equations, and the sign suggests either that older vehicles are more difficult to repair or that what must be considered the emission level for a fully repaired vehicle increases slowly with vehicle age. On average, a year of age increases post-repair emissions by 0.1 g/mi. for HC and 1 g/mi. for CO. Again, these effects are relatively trivial compared to the variation in individual vehicles.

The Sun Co. data (Table A5) show even higher pollutant removal efficiencies, owing most likely to the greater repair expenditure. Repair cost was insignificant and has been omitted from the model in the table.

The coefficients of the linear model may be biased because the dependent variable is truncated at zero, and a more searching analysis of the determinants of repair effectiveness would consider this. Such an analysis is beyond the scope of this report, and in any event would be more usefully carried out with data on the specific repairs performed. In the simulation model described in the next section we use the linear model because its coefficients are easier to interpret and because we are more interested in prediction of the value of the dependent variable than in the coefficients of the independent variables.

The higher R-squares found in the regressions on the California data (Table A4) than found in the Sun regressions (Table A5) are misleading. If, rather than the emission levels themselves we had regressed the change in emissions (i.e. pre-repair emissions minus post-repair emissions), we would have found the reverse: higher R-squares in regressions on Sun data set. That is, because emission reductions are smaller in the California data, the post-repair emissions track more closely with original emission levels. A more reliable indicator of the quality of the prediction is the standard error in each equation, which is the same regardless of whether the dependent variable is absolute post-repair level or the difference.

Table A4
Predicting Repair Effectiveness:
California I/M Review

	HC	CO	NOx
Constant	0.041 (0.25)	5.30 (1.40)	0.14 (0.048)
HC_0	0.36 (0.027)	0.53 (0.15)	-0.0025 (0.0051)
CO_0	0.026 (0.0045)	0.62 (0.024)	0.0027 (0.00084)
NOx_0	0.32 (0.13)	0.60 (0.72)	0.73 (0.025)
Cost	-0.00084 (0.00064)	-0.011 (0.0034)	-0.00027 (0.00012)
Std. error	2.87	15.54	0.54
R-square	0.34	0.59	0.57
n	669	669	669

Table A5
Predicting Repair Effectiveness:
Sun Oil Company

	HC	CO	NOx
Constant	0.36 (0.28)	2.48 (3.45)	1.04 (0.24)
HC_0	0.098 (0.026)	0.16 (0.32)	0.10 (0.023)
CO_0	-0.0036 (0.0015)	-0.00051 (0.018)	-0.0041 (0.0013)
NOx_0	0.0055 (0.020)	-0.26 (0.24)	0.011 (0.017)
Cost	0.0013 (0.00075)	0.023 (0.009)	0.00017 (0.00066)
R-square	0.11	0.05	0.12
Std. error	1.20	14.58	1.05
n	151	151	151

RESTORATION-BASED MEASURES OF COMPENSATION IN NATURAL RESOURCE LIABILITY STATUTES*

Carol A. Jones,
Damage Assessment Center,
National Oceanic & Atmospheric Administration (NOAA)

and

Katherine A. Pease, Esq.,
Office of General Counsel,
National Oceanic & Atmospheric Administration (NOAA)

Abstract:

In the past two decades, the U.S. Congress passed several major environmental statutes that designate natural resource management agencies as trustees of the resources on behalf of the public, and allow the trustees to recover damages for injuries to public resources from releases of hazardous substances and discharges of oil. The primary federal statutes are the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA, more commonly known as Superfund) and the Oil Pollution Act (OPA) of 1990. [When OPA was promulgated in 1990, its natural resource liability provisions for oil spills superseded those previously established in the Clean Water Act in 1978.]

The standard measure of damages in the various statutes is the cost of restoring the resources to baseline (*but-for* the spill) plus the interim loss in value from the time of the incident until full recovery. However, trustees are allowed to spend their recoveries ***only*** on enhancing or creating ("restoring, rehabilitating, replacing or acquiring the equivalent of") natural resources. The statutory restriction on the use of the recoveries has motivated the development of an alternative to a monetary measure of interim losses: compensation in the form of resource projects, or "resource compensation".

As a result of a wide-ranging public dialogue during the process of promulgating regulations for the resource liability provisions of OPA, NOAA incorporated the concept of resource compensation in the final regulations published on January 5, 1996. Further, some of the bills to reauthorize CERCLA currently under consideration in Congress would incorporate these concepts in statutory language.

The reframing of the damage claim poses some advantages, as well as some methodological challenges. To provide the context for our discussion of methodological approaches to measuring resource compensation, we first identify the statutory measure of damages, and the traditional framing of the valuation of injuries. We then motivate the revised measure of compensation and identify alternative methodological approaches that may be employed to implement it.

version: January 29, 1996

* Carol A. Jones, Ph.D., is Chief of the Resource Valuation Branch, Damage Assessment Center, National Oceanic & Atmospheric Administration (NOAA), Silver Spring, MD; Katherine A. Pease, Esq., is Senior Counselor for Natural Resources, Office of General Counsel, NOAA, Long Beach, CA. The views expressed in this paper are those of the authors and do not necessarily reflect the views of NOAA.

In the U.S., the atmosphere, oceans, estuaries, rivers, and plant and animal species are public trust resources. For the most part, the U.S. has not created private ownership rights to these resources but instead has allowed the public free access, while developing a system of public management to promote beneficial uses. In the past two decades, public policies have emphasized protecting the resources from injury and depletion. In particular, the U.S. Congress passed several major environmental statutes in the 1970s that contain provisions designating the management agencies as trustees of the natural resources on behalf of the public, and allowing the trustees to recover damages for injuries to public resources from releases of hazardous substances and discharges of oil.

The primary federal statutes[1] containing provisions establishing liability for injuries to resources in the public trust are the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA, more commonly known as Superfund) and the Oil Pollution Act (OPA).[2] [When OPA was promulgated in 1990, its natural resource liability provisions for oil spills superseded those previously established in the Clean Water Act in 1978.] These Acts call on the President and State governors to designate officials to serve as trustees for natural resources on behalf of the public. Natural resources are defined broadly to include land, fish, wildlife, biota, air, water, ground water, drinking water supplies, and other such resources belonging to, managed by, held in trust by, appertaining to, or otherwise controlled by the United States, any state or Indian tribe or any foreign Government.

---

[1] Some states have similar statutes or may rely on other legal theories to recover for injury to natural resources.

[2] Other federal statutes containing natural resource trustee provisions include the National Marine Sanctuaries Act (formerly the Marine Protection, Research, and Sanctuaries Act); the Federal Water Pollution Control Act (*or* Clean Water Act),; Deepwater Port Act of 1974; Outer Continental Shelf Lands Act Amendment of 1978; and Trans-Alaska Pipeline Authorization Act.

Trustees are allowed to spend their recoveries ***only*** on enhancing or creating ("restoring, rehabilitating, replacing or acquiring the equivalent of") natural resources. To ensure the public role in restoring injured natural resources, private causes of action for damages caused by injury to public trust resources generally have been narrowly defined.[3] In 1986 and 1987, the US Department of the Interior promulgated regulations for natural resource damage assessments (NRDA) under CERCLA.[4] The National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce promulgated NRDA regulations for OPA on January 5, 1996.[5] Resource valuation issues have been the subject of intensive public scrutiny during the OPA rule-making process, (as they continue to be in the CERCLA reauthorization process). During the rule-making process, academic economists joined with the commercial and environmental interest groups in submitting extensive public comments on valuation issues. As a result of the wide-ranging public dialogue, NOAA re-framed the concept of compensation in damage claims to place greater emphasis on restoration of public resources. Some of the bills to reauthorize CERCLA currently under consideration in Congress would incorporate these concepts in statutory language.

The reframing of the damage claim poses some advantages, as well as some methodological challenges. To provide the context for our discussion of methodological approaches to measuring resource compensation, we first identify the statutory measure of damages, and the traditional framing of the valuation of injuries. We then motivate the revised measure of compensation and identify alternative methodological approaches that may be employed to implement it.

[3] See Jones, Carol, Theodore Tomasi, and Stephanie Fluke, "Public and Private Claims in Natural Resource Damage Assessments," *Harvard Environmental Law Review*, Vol. 20, Number 1:111-163, 1996.

[4] The agency is currently subject to a second round of suits challenging the regulations.

[5] 61 Fed. Reg. 440 (January 5, 1996).

MEASURES OF DAMAGES

The several statutory liability provisions are based on the common law principles of the public trust doctrine[6] and parens patriae[7] whereby the sovereign has certain legal obligations to protect and preserve the trust corpus.[8] At the same time, they expand the measure of damages beyond traditional common law principles. Generally, under common law, the measure of damages for injury to natural resources or property was the diminution of value as a result of that injury.[9] However, courts in the twentieth century increasingly recognized that diminution in value was not always adequate to make the claimant "whole." In those cases, the courts awarded restoration costs.[10] The public liability statutes cited above codify this trend.

Natural resource damage claims under the various natural resource liability statutes have three basic components. The most specific language defining the components of the claim appears in the most recent statutes. For example, Oil Pollution Act specifies the measure of damages[11] as:

---

[6] The public trust doctrine provides that the government hold in trust property and natural resources for the benefit of the public. *See* Ward & Duffield, *Natural Resource Damages: Law & Economics* (1992) at 11-21 for a discussion of the doctrine.

[7] *Parens patriae* is similar to the public trust doctrine and provides the legal basis for a state to assert a claim on behalf of its citizens when their health or welfare is threatened. *Id.* at 21-23.

[8] *See, e.g., Sierra Club v. Department of the Interior*, 376 F. Supp. 90 (1974), 398 F. Supp. 284 (N.D. Cal. 1975).

[9] *See* Annot., 69 A.L.R.2d 1335 (1960).

[10] *See, e.g., Feather River Lumber Company v. United States*, 30 F.2d 642 (9th Cir. 1929)(awarding cost of restoring young forest destroyed by fire on public land); *Heninger v. Dunn*, 101 Cal. App.3d 858, 162 Cal. Rptr. 104 (1980) (suggesting that claimant is entitled to restoration of his injured land). *See also* discussion in *Ohio v. Department of Interior*, 880 F.2d 432, 455 n37 (D.C. Cir. 1989).

[11] 33 U.S.C. § 2706(d). See also the measure of damages for the National Marine Sanctuaries Act, 16 U.S.C. § 1432(6); the Clean Water Act 33 U.S.C. § 1321 (f)(4) & (5); and CERCLA 42 U.S.C. § 9607(f)(1). See *Ohio v. the Department of the Interior,* the D.C. Circuit Court opinion in the litigation over the regulations implementing CERCLA, 880 F.2d at 454 n.34., 458, 464, 476-478.

- the cost of restoring, rehabilitating, replacing, or acquiring the equivalent of, the damaged natural resources ("primary restoration");
- the diminution in value of those natural resources pending recovery of the resource to baseline, but-for the injury ("interim lost value"); and
- the reasonable cost of assessing those damages.

In this paper, we focus on the first two components of claims.

In cases of releases of hazardous materials that do not readily degrade in the environment, such as heavy metals, acid mine runoff, or PCBs, natural recovery may not occur,[12] and active human restoration may be necessary to accomplish recovery of the resources to baseline. In other circumstances, the feasible options will include natural recovery, as well as active human restoration options such as on-site restoration or rehabilitation, or off-site replacement and/or acquisition of equivalent resources.

Primary restoration projects may both expedite and increase the likelihood of recovery; consequently, the benefits of projects are reductions in the interim loss to the public of the injured resources. For example, mining contamination of the surrounding watershed may expose fisheries and forests to toxic run-off, impairing commercial harvests of fish and timber, as well as hiking, fishing, and other recreational opportunities in the affected area.

Figure 1 illustrates the relationship between primary restoration and interim losses. Time is represented on the horizontal axis. The vertical axis represents the value of services provided by an ecosystem affected by a particular release occurring at time t_0, say a spill of #6 fuel oil that oils a tidal wetland area. The oiling causes a die-back in the wetland vegetation, in addition to exposing various animals to oil, thereby impairing various services provided by the wetland.[13]

[12] That is, natural recovery may not occur within a human timeframe, such as a generation, or possibly even a human lifetime.

[13] For example, on-site ecological services may include faunal food and shelter, sediment stabilization, nutrient cycling, and primary productivity. Off-site human services, supported by the on-site ecological

Relationship between Restoration and Interim Lost Value

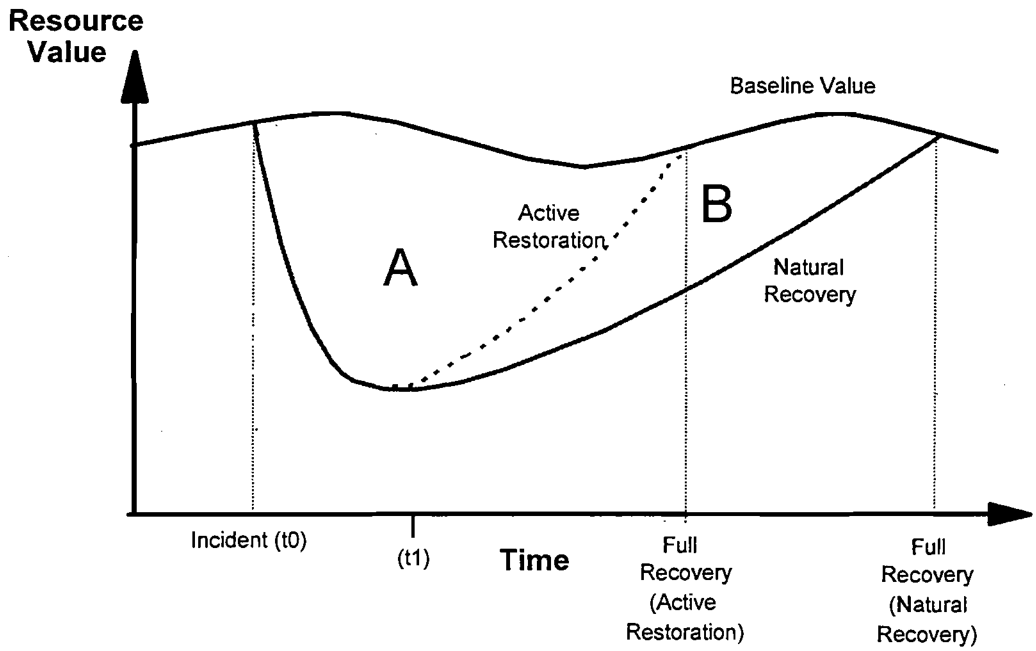


Figure 1

Assume natural recovery is feasible, as well as one "active" primary restoration technique -- re-vegetating the wetland. As illustrated in the figure, active restoration expedites the return of wetland services to baseline. Thus, the interim lost value associated with the natural recovery scenario (area A+B) is higher than for the active restoration alternative (area A only).

---

functions, may include water quality improvements, storm protection and flood control for shoreline properties, as well as bird watching along the flyway, and commercial and/or recreational fishing.

To determine if the benefits of the active project are in proportion to the costs, the restoration costs (not illustrated in the diagram) can be compared against the benefits of the project, area B. The analysis of project benefits should include an assessment of the likelihood of success as well as any external environmental or public health impacts, as well as the opportunity costs of using public resources for this project relative to alternative uses, (none of which is represented in the diagram). Note however that, in the litigation over the NRDA regulations implementing CERCLA, the court ruled out a strict benefit-cost test. Rather it determined that the statute embodied a distinct preference for restoring the resources to baseline, except where restoration is "practically infeasible" or where the cost of restoration is "grossly disproportionate" to the value of the resource.[14]

In the next section we provide the conceptual economic framework for a fully restoration-based measure of compensation for injuries. In the following section, we outline several methodological approaches for calculating this measure of damages.

CONCEPTUAL FRAMEWORK FOR COMPENSATION

The key difference between the fully restoration-based measure of damages and prior approaches is in the treatment of interim losses. To highlight the differences, we first review the conventional economic measure of damages for interim losses.

Given the premise that the public holds the property rights to public resources, economic theory implies that the correct framework for compensation for injury to the resources is willingness-to-accept. The measure of monetary compensation, then, is the minimum amount of money necessary to make individual members of the public as well off as they would have been *but-for* the discharge. (The measure is conditional upon the rate of recovery of public resources

[14] See *Ohio v. Department of Interior*, 880 F.2d at 443 n.7, 456, 459.

to baseline levels, accomplished either through natural recovery or active human restoration.) In litigation where individual claimants receive monetary compensation, for example in private tort suits, the monetary compensation measure provides the damages necessary to "make whole" individual claimants.

To develop the analytical framework for characterizing monetary compensation, we use a lifetime utility function to accommodate the intertemporal nature of injuries and restoration projects:

$$U = U(\boldsymbol{P}, \boldsymbol{Y}; \boldsymbol{Q})$$

where U represents an individual's baseline level of lifetime utility, P is the price vector through time for private goods, Y is the individual's lifetime wealth (reported in present discounted terms, assuming full borrowing and lending at real discount rate r), and $\boldsymbol{Q}$ is the vector representing the baseline flow of services provided by public resources through time.

We represent post-injury service flows from public resources, assuming no active human restoration processes, as $\boldsymbol{Q'}$, where some of the resources represented in vector $\boldsymbol{Q'}$ are impaired for a period of time due to the injury, and others are not impaired.[15] As noted above, depending upon the characteristics of the spill or release, natural recovery may or may not occur. Given the legislative mandate to restore injured resources to their baseline, the vector $\boldsymbol{P^k}$ represents the augmentation of the service flows provided by primary restoration project k, where projects $(k=1,\ldots,K)$ are under consideration.

[15] We assume that price and incomes are unchanged as a result of the release, simply for the sake of notational simplicity. It would be straightforward to incorporate the possibility that prices and incomes may change as a result of the injury, as a result of primary restoration.

We can then characterize for an individual the willingness-to-accept measure of monetary compensation, M^k, for the combined effects of resource impairment from $\boldsymbol{Q}$ to $\boldsymbol{Q'}$ and primary restoration project $\boldsymbol{P^k}$:

$$(1) \qquad U = U(\boldsymbol{P}, \boldsymbol{Y}; \boldsymbol{Q}) = U(\boldsymbol{P}, Y + M^k; \boldsymbol{Q'} + \boldsymbol{P^k}) \qquad \text{where } \boldsymbol{Q'} < \boldsymbol{Q},$$

where the prime superscript refers to the post-injury state, and monetary compensation M^k represents compensation in present discounted value terms (assuming all is paid in the current period).[16]

Note that in Figure 1, M^k (conditional upon the choice of primary restoration project, $\boldsymbol{P^k}$) is illustrated in its undiscounted form by areas $A+B$ and A, for natural recovery and active restoration projects, respectively.

The measure of damages then is the cost of primary restoration project $F_P(\boldsymbol{P^k})$ plus the sum across all individuals, $n=1,\ldots,N$, of monetary compensation for the interim loss in resources:

$$(2)\ Damages = F_P(\boldsymbol{P^k}) + S_n M^k{}_n$$

As noted above, public trustees do not have the authority to make individuals whole by providing such recoveries directly to individuals; rather, trustees are allowed to spend their recoveries ***only*** on enhancing or creating ("restoring, rehabilitating, replacing or acquiring the equivalent of") natural resources. Note that, if the compensation *were* paid to individuals in the form of money, there would be analogous restrictions on the use of the compensation monies; the recipients of private recoveries essentially would be limited to spending the monies on private goods. Individuals generally cannot make autonomous decisions regarding expenditures on public resources: public decision-making processes must be invoked.

[16] For notational simplicity, we assumed the injury did not affect prices and incomes; that assumption could readily be relaxed.

The statutory restrictions on the use of the recoveries have motivated the development of an alternative measure -- resource compensation. We represent a compensatory restoration project as C^k, which combined with primary restoration project P^k, forms the restoration alternative, R^k. The superscript k on the compensatory project denotes that it provides sufficient compensation for interim losses, if primary restoration project P^k is performed. Compensatory projects may be either off-site habitat creation or enhancement or on-site restoration beyond pre-spill baseline (e.g., at sites for which the baseline conditions are degraded relative to their potential).[17]

Note that public resources are to be shared in common with all members of the public. Consequently, resource compensation cannot be calculated by replacing individualized monetary compensation amounts in equation (1) with individualized service levels provided by a public resource project, and then simply summing services across individuals. Rather we characterize resource compensation as the scale of resource projects for which the sum of net surplus gains (losses) for the injury and the project across individuals equals zero.

To this end, we introduce the net-change-in-surplus measure, $W^k = W(Q', P^k, C^k)$, which represents the amount of additional income an individual would pay $(W^k_n > 0)$ [would require,

$(W^k_n < 0)$] for the combined effects of the resource injury and the provision of restoration alternative R^k, comprised of primary and compensatory restoration projects (P^k, C^k). Applying the notation developed above, we characterize the "net change in surplus", W^k, for an individual as follows:

---

[17] Note that though the concepts of *primary restoration* and *compensatory restoration* are distinct, the primary and compensatory restoration projects may involve the same restoration activity (e.g., planting the seagrass *Syringodium* at a site where the seagrass has been destroyed by a grounding) on contiguous sites. Alternatively, the projects may involve different restoration activities (e.g., natural recovery at a wetland injury site, and wetland creation at the compensatory site) at different locations within the same estuary.

$$(3) \qquad U = U(P, Y; Q) = U(P, Y + W^{k};\ Q' + P^{k} + C^{k}), \qquad \textit{where } Q' < Q,$$

To determine which projects satisfy the requirements of providing resource compensation to the public, we employ a Hicks-Kaldor compensation framework in which gainers could compensate losers for the effects attributable to the combination of the injury and the resource project. The set of resource projects R^k that provide "equivalent resources and services" [18] to the injured resources -- resource compensation -- is the set of projects $\boldsymbol{R^k}$ for which:

$$(4) \qquad S_n W^{k}_{n} = 0$$

where we now write out the previously implied subscript n to represent the summation over all members of the public $(n=1,\ldots,N)$ of the net change in surplus measure. Under a resource compensation framework, the measure of damages then becomes:

$$(5) \qquad Damages = F_P(\boldsymbol{P^k}) + F_C(C^k)$$

where C^k is the minimum cost resource project of the set of compensatory resource projects that provide "equivalent resources and services" to the injured resources.

There is no necessary relationship between the costs of resource compensation $F_C(C^k)$ and monetary compensation, $S_n M^{k}_{n}$, for interim losses:[19]

$$(6)\ F_C(C^k) \geq S_n M^{k}_{n} \qquad or \qquad F_C(C^k) \leq S_n M^{k}_{n}$$

[18] For example, to compensate for recreational beach losses, alternative projects might include building boardwalk over sand dunes (to provide access to the beach while at the same time protecting, and providing access to, the fragile dune habitat); or constructing near-shore artificial reefs for snorkeling, diving, or fishing.

[19] This formulation assumes that the cost of primary and compensatory restoration are independent. More generally: $\Phi(C,P) \geq \Phi(P) + M$ **or** $\Phi(C,P) \leq \Phi(P) + M$.

The costs of compensatory restoration projects may be greater than, equal to, or less than monetary compensation. The potential divergence may result from several effects. First of all, natural resources are not produced in a market - consequently, there are no market incentives for producers to adjust the quantity produced so that the value of services and the cost of producing the services move toward equivalency on the margin. Consequently, the baseline supply of public resources may be in excess demand [excess supply] rather than at equilibrium, with the result that value may be greater than [less than] supply costs at the baseline level of resources.

If the impacts of the injury and the replacement project on total quantity supplied (and consequently, on value and cost per unit) are marginal, then divergent pre-spill value-cost relationships will carry over to the impacts of the injury and restoration projects. With marginal impacts on the supply of resources, the minimum quantity of discounted resource-years necessary to compensate for the injury may equal the discounted quantity of comparable resource-years that were lost. Nonetheless, because of the initial disequilibrium, the cost of compensatory restoration $[F_C (C^k)]$ may be less than [greater than] the lost value $[S_n M^k_n]$.

Second, even if there were a market process resulting in a baseline equilibrium state, both the injury and the replacement project may be sufficiently large that the effects of each are not on the margin. The reduction in supply from the injury at the site could result in a higher value per unit of services lost than at equilibrium (with value greater than cost). Further, if most of the replacement benefits occur after the injury site has recovered from the injury, then the compensatory increase in supply could be valued at less than the equilibrium value [with value less than cost]. In this case, the minimum quantity of discounted resource-years necessary to compensate for the injury may exceed the quantity of discounted resource-years that were lost.[20] Further, the cost of restoration per unit will exceed the value per unit. Both factors reinforce the possibility that the costs of resource compensation may exceed monetary compensation.

Consider the following hypothetical example illustrated in Figure 2. The closure of half of the hiking trails in a wildlife refuge due to contamination may cause a substantial loss in use during the period of closure. However, after the site is decontaminated and all original trails reopened, further increasing hiking trails to compensate for the lost trail mile-years[21] in fact may not add much additional value, if the demand for hiking was fully satisfied by the original quantity of trails. *DD'* is an annual demand function for hiking trail miles at the site. Demand is basically sated at the baseline supply of trails providing a capacity for T_0 trips, so the marginal trip is of value, e. Assume that, due to hazardous contamination, the supply is reduced for a period to a capacity of $T_0 - J$ trips. At that level of supply, the value of a marginal trip is $v>e$, and the lost consumer surplus is the area A.[22] When the previously restricted trails are decontaminated and reopened, if additional trails were to be added that provided capacity for $T_0 + R$ trips, the increment in consumer surplus would be minimal.

However, the foregoing discussion is very simplistic, because it completely ignores the question regarding how to characterize "services" and the closely related question as to the

---

[20] V. Kerry Smith makes this point in his paper, "Natural Resource Damage Liability: Lessons from Implementation and Impacts on Incentives", January 1994, p. 9, citing the work of Carson, R., N. Flores, and W. M. Hanemann, "On the nature of compensable value in a natural resource damage assessment", presented at the AEA meeting, January 1992.

[21] The replacement equation can be solved by assuming a finite lifetime for the additional trail miles and discounting the hiking benefits in the future and compounding the losses from the past.

[22] This is the correct calculation of consumer surplus only when the supply is rationed by setting the entry fee high enough so that supply just equals demand. Alternatively, if supply is rationed on a first-come, first-served basis, then the lost consumer surplus per mile will be the average consumer surplus under the demand function.

Demand for Wildlife Refuge Hiking Trips

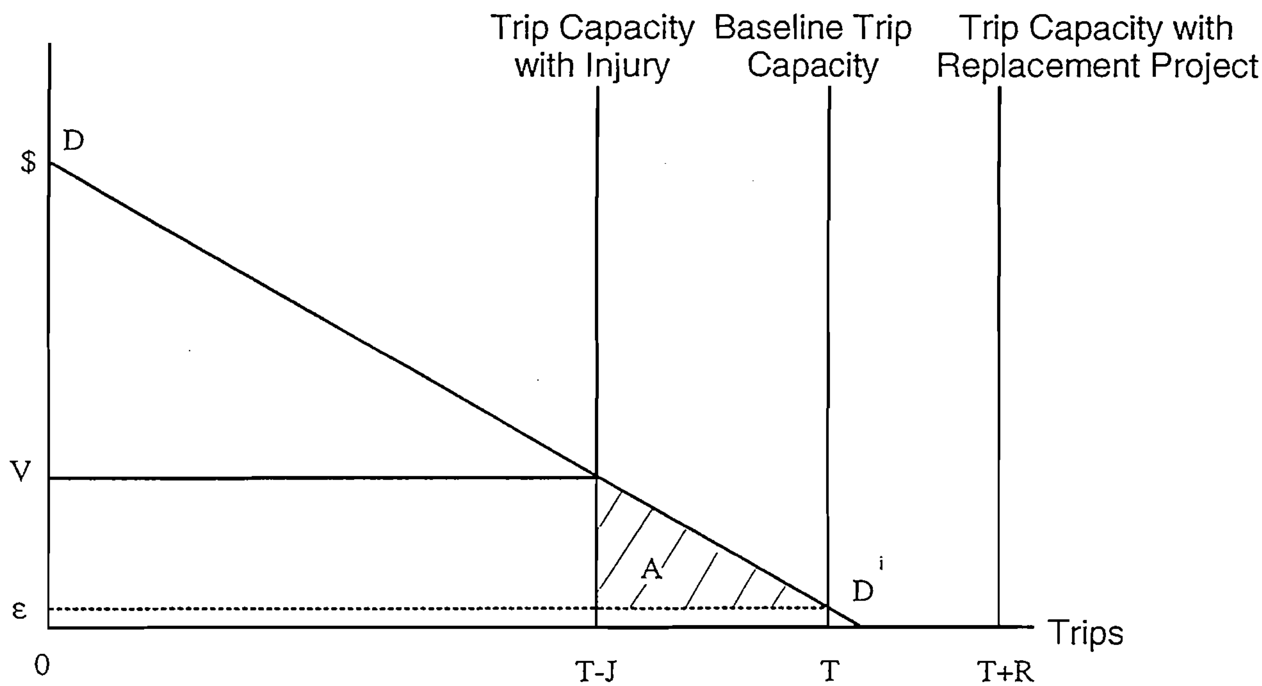


Figure 2

"scope" of the market for resource services. A critical question is, what is the range of resource projects that are considered substitutes for the lost services and resources? The more broadly the allowable activities are defined, the more likely it will be to find one that provides value at a cost that is less than the value provided.

The statutes restrict the scope of projects that may be considered, with the mandate to "restore ... the equivalent" of the injured resources. Consequently, when selecting compensatory restoration projects, trustees must demonstrate a nexus between the injured resources and lost services and the resources and services provided by the replacement project. However, even when replacement projects involve the same habitat as the injury site, substitutions in resources and/or geographical locations are unavoidable to some extent. Consider the following examples:

- A salmon stream previously supporting wild salmon stock is contaminated and supports no salmon for 30 years. If the metric for services is defined as salmon returning to the stream to spawn, with no reference to the genetic stock of the salmon eliminated from the stream, will the quality of the resource lost be achieved in a replacement project using hatchery fish?
- The habitat injured by a hazardous release is hardwood bottom land (or mangrove, or *Thalassia* seagrass) wetlands. In order to recreate such a habitat, the site must first progress through several other habitats before achieving that stage in habitat succession. How should the metric capture the habitat substitution during the succession process?

We turn now to consider the procedures that are available to estimate the amount of resources necessary to provide resource compensation.

PROCEDURES FOR CALCULATING RESOURCE COMPENSATION

After the interim loss in resources due to an injury has been quantified, calculating resource compensation for the interim loss requires several steps:

- identifying restoration alternatives;
- determining the scale of restoration alternatives that would compensate for the interim losses, and then

- selecting the most preferred restoration alternative, based on cost-effectiveness and other relevant criteria.

Focusing on the second step, we consider three possible approaches to determining the scale of projects required to compensate the public.

1. Stated Preference Methods Eliciting Direct Resource-to-Resource Trade-offs

With stated preference methods, it is possible to elicit direct tradeoffs between resource injuries and resource increments provided by compensatory restoration projects. Some researchers have suggested that contingent choice analysis,[23] which elicits from survey respondents their choices among alternative "product" scenarios with varying quality, quantity, and cost attributes, may be suited to determining resource compensation.[24] Contingent choice analysis has been extensively used to assist firms in designing new products, a problem with similarities to the design of restoration plans. Applications of conjoint analysis to environmental or health risk policy issues are beginning to appear in the literature.[25]

The method allows the analyst to map out valuation surfaces over reductions and increments in quality and quantity attributes of goods or services, (which could be resource

[23]For a survey of the literature see, Jordan J. Louviere, "Conjoint Analysis Modeling of Stated Preferences," *Journal of Transportation Economics and Policy*, January 1988.

[24] See for example, Mazzotta, M., James J. Opaluch, and Thomas A Grigalunas, "Natural Resource Damage Assessment: The Role of Resource Restoration," *Natural Resources Journal*, Vol. 34, No. 1, Winter 1994 (153-178).

[25] See, for example, Christopher Gan, and E. Jane Luzar, "A Conjoint Analysis of Waterfowl Hunting in Louisiana," *Journal of Agricultural and Applied Economics*, 25:36-45, 1993; Alan Krupnick and Maureen L. Cropper, "The Effects of Information on Health Risk Valuations," *Journal of Risk and Uncertainty*, 5:29-48, 1992; John Mackenzie, "A Comparison of Contingent Preference Models," *American Journal of Agricultural Economics*, 75(3):593-603, 1993; James Opaluch et al. "Evaluating Impacts from Noxious Facilities: Including Public Preferences in Current Siting Mechanisms", *Journal of Environmental Economics and Management*, 24:41-59, 1993; W. Kip Viscusi, Wesley A. Magat, and Joel Huber, "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis," *Journal of Environmental Economics and Management*, 21:32-51, 1991; Lisa Wood, Anne E. Kenyon, William H. Desvouges, and Lyn K Morander, "How much are customers willing to pay for Improvements in Health and Environmental Quality?" *The Electricity Journal*, 70-77, May 1995.

scenarios). In the design of new consumer products, the goal is to select the profit-maximizing combination of quality, quantity, and price attributes. Contingent choice analysis is employed to derive an expected revenue function defined over different combinations of quantity, quality and price; in combination with a cost function, the profit maximizing product design (and market scale) can be determined.

In calculating resource compensation, the goal is to determine the *scale* of a restoration alternative sufficient to compensate for resource injuries. In this application, the method is employed to derive a valuation function to measure individual surplus W for the combined impacts of the injury and of restoration alternatives, and then calculate the scale and quality attributes of alternatives such that aggregate consumer surplus is zero ($SW=0$).

To elicit the information required for such a determination, the alternative scenarios offered to respondents must be carefully designed. It is critical to identify the quality, quantity (and cost) attributes that either were impaired due to the injury and/or will be provided by compensation projects, and to determine the relevant range of variation for each. A base case resource scenario is developed as a reference point, and alternative resource scenarios are designed to cover the range of relevant range of variation.[26] Notably, the specific details of feasible compensatory projects need not be pre-conceived by the survey designers and included as alternative project scenarios. Individual respondents are asked to make a series of choices among the base case and alternative resource scenarios.

A design challenge for measuring resource compensation, which does not appear to have an analog in typical market research applications, is to incorporate resource losses in the scenarios in a way that maintains the respondents' incentives to respond truthfully to the choices.

[26] Typically, the levels of the different attributes for each alternative are assigned across the set of alternatives on statistical efficiency grounds; statistical issues also affect how alternatives are selected for inclusion in each choice set posed to the respondent.

Another design challenge is to capture the critical attributes defining the services provided by resources, an issue discussed in the previous section.

For example, consider the case of shoreline and navigational waterway closures mandates to enable the Coast Guard to stabilize and remove a barge grounded near-shore that is leaking fuel oil; the closures resulting in preventing the use of recreational beaches and fisheries. Alternative projects to compensate for these lost uses might include building boardwalk over sand dunes (to provide access to the beach while at the same time protecting, and providing access to, the fragile dune habitat); constructing fishing piers (to facilitate shore-based access, in an area with little or none currently); or developing near-shore artificial reefs for snorkeling, diving, or fishing (for beach- and boat-based access). The losses due to the beach and fishing closures could be incorporated in the alternative project scenarios as closure periods of the resources that are necessary for construction of the different versions of the compensatory projects. However, such a scenario would only capture the effects of the closure per se, not any direct resource losses or continuing impairments after the opening of the closure.

2. Two-part Calculation: Losses and Gains

Alternatively, the compensatory restoration scaling process can be decomposed into a two-part valuation process, in which losses due to injury and gains from projects are valued separately. We propose the following decomposition of W^k into two components, where the first is the monetary measure of damages for the interim loss of resources, and the second is the surplus generated by the provision of the compensatory restoration project R:

$$(7) \qquad W^k = M(P, Y, Q' + P^k; P, Y, Q) - S(P, Y, Q' + P^k + C^k; P, Y, Q)$$

Based on *(7)*, the requirement that $SW^k = 0$ is equivalent to the requirement that the sum of the losses equals the sum of the gains:

$$(8) \qquad S\,M(P,\ Y,\ Q' + P^k;\ Q) = S\,S(P,\ Y,\ Q' + P^k + C^k;\ Q)$$

A single survey eliciting stated preferences may be employed, in order to capture the total value of the losses and the gains. Alternatively, for recreational losses, a travel cost study may be suitable. If the environmental characteristics of the injuries and/or the restoration projects are outside the range of currently observed circumstances, then linking contingent behavior data with the observed participation data may be required to complete the travel cost analysis. Combining stated preference and revealed preference data provides additional advantages, including allowing identification of attribute effects that cannot be identified with revealed preference data due to collinearity in the dataset. Collinearity can be reduced in the combined dataset by judicious design of contingent scenarios.[27]

In some situations, it may be more cost-effective to perform independent calculations of gains and losses, particularly if suitable benefits transfer studies are available. If the compensatory project will not begin to provide resource services until after the injured resource has returned to its baseline level (Q), then we may rewrite *(8)*:[28]

$$(8') \qquad S\,M(P,\ Y,\ Q' + P^k;\ Q) = S\,S(P,\ Y,\ Q + C^k;\ Q)$$

[27] Adamowicz, W., J. Louviere, and M. Williams, "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management* 26, 271-292 (1994).

[28] Alternatively, if the compensatory projects begin to provide services before full recovery, *(7')* will be an approximation of the calculation in equation *(3)* unless the utility function is linear in the resource service levels over the relevant range

If we assume a small injury, for example a small injury to a habitat resource, then we can perform a Taylor expansion of *(8')* around baseline levels of ***Q*** and ***C***, to yield the following approximation:

$$(9) \qquad S_n\, S_t\, v_{nt}(\mathit{loss\ in\ services/\ acre})_t\,/(1+r)^t * J\ \mathit{acres\ of\ habitat\ injured} \quad =$$

$$S_n\, S_t\, w_{nt}(\mathit{gain\ in\ services/\ acre})_t\,/(1+r)^t * R\ \mathit{acres\ of\ replacement\ habitat,}$$

where v_{nt}, refers to the value per acre of injured habitat to individual n in period t, w_{nt} refers to the value per acre of replacement habitat to individual n in period t, and r is the real rate of discount. The equation is then solved for the scale of the compensatory restoration project, C acres of habitat.[29]

For conducting expedited analyses, there is a paucity of existing data suitable for a benefits transfer-type valuation of habitats. Consider for example, a wetland habitat with limited direct human use on-site. Ecological services on-site may include faunal food and shelter, sediment stabilization, nutrient cycling, and primary productivity. Off-site human services, supported by the on-site ecological functions, may include water quality improvements due to on-site water filtration, storm protection and flood control for on-shore properties, as well as bird watching along the flyway and commercial and/or recreational fishing. These off-site services are difficult to quantify even with comprehensive site-specific studies. Further, in part for the same reasons, it is difficult to estimate the impairment of these services as a result of spills or

[29] In its essence, the method calculates the scale of projects required to replace present discounted quantities of effective habitat acres (or more generally, effective resource units). The method is closer to an asset replacement approach than a service valuation procedure. For further discussion, see "Habitat Equivalency Analysis: An Overview," Damage Assessment and Restoration Program Policy and Technical Paper Series, National Oceanic and Atmospheric Administration, March 15, 1995.

releases. Finally, we may not have good value estimates for habitats that are suitable for transfer from one site to another in a benefits-transfer type procedure.[30]

Under certain conditions,[31] the values per acre cancel out of the equation, and the equation can be solved in terms of quantities of services (i.e., percent function of an acre) lost and gained. This "service-to-service" approach is referred to as Habitat Equivalency Analysis (HEA), when a habitat is injured.

3. A Simplified Procedure: Habitat Equivalency Analysis[32]

In essence, HEA calculates the scale of projects to replace injured resources in terms of present discounted quantities of "effective" habitat acres (or more generally, "effective" resource or service units). The simplified calculations in Habitat Equivalency Analysis (HEA) are appropriate when it has been determined that the project(s) selected by the trustees will provide replacement resources and services of "same type, quality, and comparable value" to the injured resources.[33]

---

[30] Benefits will depend on various site-specific factors that may not be well-documented in the literature studies (e.g., for flood control, the retention capacity of the wetland and the frequency of serious storm events); further, existing research valuing wetland services addresses only a limited number of service flows (e.g., Farber and Costanza [1987] only addresses fisheries, storm protection).

[31] The simplest assumptions is that $v_{nt} = w_{nt} = v$.

[32] Other authors discuss the general concept, but suggest different possible criteria for appropriate applications. See Mazzotta, M., James J. Opaluch, and Thomas A Grigalunas, "Natural Resource Damage Assessment: The Role of Resource Restoration," *Natural Resources Journal*, Vol. 34, No. 1, Winter 1994 (153-178); and Unsworth, Robert E. and Richard C. Bishop, "Assessing Natural Resource Damages using Environmental Annuities," *Ecological Economics* 11(1994) 35-41.

[33] Conceptually, the method compensates for the loss of total direct use services through time due to resource injury and captures an unknown portion of passive use losses.

To meet these conditions in applying Habitat Equivalency Analysis, trustees are to select projects for which, in their judgment, the values per unit of lost services and replacement services per "effective" habitat acre are comparable.[34]

The trustees must specify a variety of inputs to the calculation in order to characterize the reduction in habitat "services" during the injury recovery process and the increment in services provided per acre of the replacement projects. A critical methodological issue in implementing HEA is to identify a metric that can be used to characterize the variations in services per acre for both habitat sites. The metric will generally be specified as a measure of on-site ecological services;[35] however, it is also being used as a proxy for the level of human direct uses and ecological services occurring off-site.

Further, trustees must make qualitative judgments that the replacement project will provide habitat services of comparable value to the services lost due to the injury. These require judgments that the chosen metric adequately captures variations in quality at the sites, and that

[34] A prior condition is that the services at the injury and replacement project sites can be characterized in a common metric, in order to define "effective" habitat acres. There are two major reasons why equivalency over discounted quantities of services may *not* provide comparable value: differences in quality, or differences in the total stock of resource services available at a given point of time (scale effects). These correspond to shifts of the demand curve vs. movements along the demand curve.

[35] For example, it may be an indicator service, which stands in for all the others, or a weighted average of multiple services. The level of passive use services, if any, are not necessarily linear in this measure. Passive use values pertain to the values that individuals place on the existence of natural resources for their own sake or for family, friends or future generations. The term 'nonuse' value has also been used to refer to the same concept.

Variables measuring the level of resource density, resource biomass, etc. may provide a basis for such a metric, but the metric for the *level of habitat services* may not be a linear function of the *resource* measure. For example, if the ecological service of primary concern is sediment stability, it would be inappropriate to characterize stability with a linear measure of root biomass, if there were a minimum level of biomass per square meter that must be achieved before any increment in stability occurs. In this case, it may be more appropriate to characterize sediment stability as an s-shaped function of root biomass, e.g.,: $q = F(r)$, where r is a resource measure, and q is habitat service levels.

the variations in scale of services through time at the sites do not substantially change the value per unit of services.

As a result of this particular concern,[36] in contexts where there are substantial human uses on-site (or human uses off-site that are feasible to quantify), it may be more appropriate to focus the calculation of resource compensation on specific services, rather than on resource units. The rate of use and the value per unit of resource may vary substantially over time and space and with variations in quality; with data on uses, standard economic methods are available to estimate the value of the uses.

NOAA has employed HEA in several natural resource damage assessment cases, covering injuries from chronic contamination, oil spills, and vessel groundings.

CONCLUSION

In the natural resource damage assessment regulations promulgated for Oil Pollution Act, NOAA has reframed the calculation of the interim loss component of damage claims to embody the statutory concept of compensating the public through resource restoration projects. The natural resource liability statutes have embodied a clear preference for trustees ensuring the restoration of injured resources to their *but-for* injury baseline. In addition, the regulations direct trustees to propose compensatory restoration projects of an appropriate scale to "make the public whole" for the interim loss in resources, and to claim from the responsible parties the costs necessary to implement the projects.

This resource compensation approach focuses the trustees from the beginning on the ultimate goal -- providing public resources as compensation to the public for interim losses due to

[36] Note that with human use, the quality issue takes on the additional dimensions associated with how human values of a site vary. For example, for a coral reef replacement project, access to the replacement

public resource injuries-- rather than performing the valuation of interim losses prior to resolution of the claim, and then planning compensatory restoration projects after resolution of claims. With this approach, the restoration of resources and the implementation of compensatory projects may occur more quickly and at lower cost.

A variety of tools exist to perform the calculations necessary to measure resource compensation. Identifying the most effective ways of applying them to particular contexts no doubt will raise interesting methodological issues.

---

project site may be inferior to the injury site. And do the concrete forms employed to provide a platform for transplantation of coral species from other sites provide comparable aesthetic value to divers?