WESTERN REGIONAL RESEARCH PUBLICATION

W-133
Benefits and Costs of Resource Policies Affecting Public and Private Land

Papers from the Annual Meeting
Miami, Florida, February 26-28, 2001
Fourteenth Interim Report
October 2001

Compiled by
Jerald J. Fletcher

Department of Agricultural & Resource Economics
Davis College of Agriculture, Forestry & Consumer Sciences
West Virginia University
Morgantown, WV 26506-6108

Table of Contents

Introduction

This volume contains the proceedings of the 2001 Technical Meeting of W-133 Western Regional Project *Benefits and Costs of Resource Policies Affecting Public and Private Land.* The meeting was held February 26-28, 2001 at the Miccosukee Resort and Conference Center in Miami, Florida.

The Miami meeting was attended by academic faculty from many W-133 participating universities in addition to researchers from other universities, federal agencies, and private consulting firms. A list of the meeting attendees with institutional identification and e-mail addresses follows.

The papers included in this volume represent the diverse nature of current research addressing the W-133 project objectives. The current objectives are to investigate 1) benefits and costs of agro-economic policies, 2) benefits transfer for groundwater quality programs, 3) valuing ecosystem management of forests and watersheds, and 4) valuing changes in recreational access. The complete program follows the list of participants. Note that not all papers presented are included in the proceedings and, in some instances, authorship and title were modified for the proceedings.

The development of any such program is a time consuming task. The turnout was excellent and comments indicated that W-133 had another successful meeting. As usual, there were amenity attributes for all to enjoy. Alligator wrestling may be a new one for W-133, however. I'd like to thank Steve Polasky for turning things over in great shape and Scott Shonkwiler and Frank Lupi for help during the meetings in Miami.

Jerald J. Fletcher
Agricultural and Resource Economics
West Virginia University
Morgantown, WV 26506-6108

October 2001

W-133 Meeting Attendees -- Miami – 2001		
Laura Baumann	lbaumann@ospr.dfg.ca.gov	CA Department of Fish & Game
Enoch Bell/PSW/USDAFS	ebell@fs.fed.us	USDA Forest Service
John Bergstrom	jbergstrom@agecon.uga.edu	University of Georgia
Matt Bingham	mfb@ter.com	Triangle Economic Research
Kate Carson	kate.carson@usafa.af.mil	US Air Force Academy
Susan Casement	sdcasement@ucdavis.edu	University of California, Davis
Patty Champ	champ@lamar.colostate.edu	Colorado State University
Earl Ekstrand	eekstrand@do.usbr.gov	US Bureau of Reclamation
Jeffrey Englin	englin@unr.edu	University of Nevado, Reno
Eric English	eric.english@noaa.gov	NOAA
Don Epp	dje2@psu.edu	Pennsylvania State University
Jerry Fletcher	jfletch@wvu.edu	West Virginia University
Ron Fleming	rfleming@ca.uky.edu	University of Kentucky
Kelly L Giraud	ffklg@uaf.edu	University of Alaska, Fairbanks
Therese Cavlovic Grijalva	tcavlovic@weber.edu	Weber State University
Steve Hampton	shampton@ospr.dfg.ca.gov	CA Department of Fish & Game
LeRoy Hansen	lhansen@ers.usda.gov	USDA/ERS
David Harpman	dharpman@do.usbr.gov	US Bureau of Reclamation
Thomas A. Heberlein	taheberl@facstaff.wisc.edu	University of Wisconsin
Daniel Hellerstein	danielh@ers.usda.gov	USDA/ERS
John Hoehn	hoehn@msu.edu	Michigan State University
Sandra Hoffman	sahoffmann@facstaff.wisc.edu	University of Wisconsin
Paul Jakus	pmjakus@agmail.ag.utk.edu	University of Tennessee, Knoxville
Reed Johnson	rjohnson@ter.com	Triangle Economic Research
Michael Kaplowitz	kaplowit@msu.edu	Michigan State University
Jason Kinnell	jck@ter.com	Triangle Economic Research
Linda Langner	llangner@fs.fed.us	USDA Forest Service
Douglas Larson	larson@primal.ucdavis.edu	University of California, Davis
Dan Lew	lew@primal.ucdavis.edu	University of California, Davis
John B. Loomis	jloomis@ceres.agsci.colostate.edu	Colorado State University
Frank Lupi	lupi@pilot.msu.edu	Michigan State University
Doug MacNair	dmacnair@ter.com	Triangle Economic Research
Don McLeod	dmcleod@uwyo.edu	University of Wyoming
Klaus Moeltner	moeltner@unr.edu	University of Nevada, Reno
Jim Opaluch	jimo@uri.com	University of Rhode Island
Joan Poor	joan_poor@hotmail.com	
Rich Ready	rready@psu.edu	Pennsylvania State University
Randy Rosenberger	rrosenbe@wvu.edu	West Virginia University
W. Douglass Shaw	wdshaw@unr.edu	University of Nevada, Reno
Scott Shonkwiler	jss@unr.edu	University of Nevada, Reno
Steve Shultz	sshultz@ndsuext.ondak.edu	North Dakota State University
Juha Siikamaki	juha@primal.ucdavis.edu	University of California, Davis
Tom Stevens	tstevens@resecon.umass.edu	University of Massachusetts
Branka Turcin	fsbt1@uaf.edu	University of Alaska, Fairbanks
Phil Wandschneider	pwandsch@nail.wsu.edu	Washington State University
Matthew A. Wilson	mawilson@wam.umd.edu	University of Maryland

PROGRAM
W-133 ANNUAL MEETING, MIAMI, FL, FEB. 26-28, 2001
(Presenter's name in *italics*)

Monday	7:30-8:15 Continental Breakfast	
8:15-9:45	Moderator - Paul Jakus	
	Too Many "NOs"? An Experimental Investigation of Behavior in Double-Bounded Referenda	
		A.C. Burton, Leeds University; ***Katherine S. Carson,*** United States Air Force Academy; Susan M. Chilton, University of Newcastle upon Tyne; and W.G. Hutchinson, Queens University, Belfast
	Testing the Validity of Contingent Behavior Trip Responses	
		Therese C. Grijalva, Weber State University; Robert P. Berrens and Alok K. Bohara, University of New Mexico; and W. Douglass Shaw, University of Nevada, Reno
	Is Actual Willingness to Pay Sensitive to Response Format? -- Implications for Contingent Values	
		Patricia A. Champ, USDA Forest Service, and Richard C. Bishop, University of Wisconsin
9:45-10:15	BREAK	
10:15-11:45	Moderator - Douglas Shaw	
	An Experimental Economics Test of Factors Encouraging the Voluntary Provision of Public Goods	
		Donald J. Epp and John D. Wicinas, The Pennsylvania State University
	Testing a Computer-Assisted Valuation Panel Approach for Valuing Watershed Ecosystem Restoration	
		John Bergstrom, Tom Holmes, Eric Huszar and Susan Kask, University of Georgia
	Geographical Respondent Nesting and Option, Bequest, Existence and Warm Glow Values: Case Study of Alaskan Steller Sea Lion	
		Branka Turcin and Kelly Giraud , University of Alaska, Fairbanks
11:45-1:15	LUNCH	
1:15-2:45	Moderator - Jeff Englin	
	The Infeasibility of Wetland Restoration to Reduce Flood Damage in the Red River Valley: Would the Inclusion of Non-Market Wetland Values Make a Difference?	
		Steve Shultz and Jay Leitch, North Dakota State University
	Testing the Validity of Benefit Transfers: A Site Correspondence Model	
		Randall S. Rosenberger, West Virginia University
	Experiments in Valuing Wetland Ecosystems	
		John Hoehn, Frank Lupi, and ***Michael Kaplowitz***, Michigan State University
2:45-3:15	BREAK	
3:15-4:45	Moderator - Doug Larson	
	When More Is Less: Student Valuation of Global Warming Policies	
		S Johnson and ***P Wandschneider***, Washington State University
	An Alternative Method for Combining Revealed and Stated Preference Data: Measuring the Effect of Fish Consumption Advisories on Recreational Fishing	
		Doug MacNair, Triangle Economic Research.
	What is the value of a bird? How resource equivalency analysis (REA) may be used in natural resource damage assessments (NRDA)?	
		Steve Hampton, California Department of Fish and Game

Tuesday	7:30-8:15 Continental Breakfast
8:15-9:45	Moderator - Phil Wandschneider
	Melding Bushels of Apples with Lugs of Oranges
	LeRoy Hansen and Roger Claassen
	Benefits of Natural River Flows Due to Hydropower Dam Relicensing in Michigan
	Michael Moore, University of Michigan; ***Frank Lupi***, Michigan State University; and Edward Rutherford and Matthew Kotchen, University of Michigan
	The Public's Willingness to Pay for Improving California's Water Quality
	Douglas M. Larson, ***Daniel K. Lew***, and Yuko Onozaka, University of California, Davis
9:45-10:15	BREAK
10:15-11:45	Moderator – Randy Rosenberger
	Valuing Ancient Forests: A Case Study of Jasper National Park
	Jeff Englin and Jered McDonald, University of Nevada, Reno
	Measuring Individual-Specific Shadow Values of Leisure Activities: Some Preliminary Results
	Douglas M. Larson and Daniel K. Lew, University of California, Davis
	Valuing Open Space Attributes in Colorado: Evidence from a Hedonic Analysis of Market Transactions
	John Loomis, Vicki Rameker and Andy Seidl, Colorado State University
11:45-1:15	LUNCH
1:15-3:00	Land Use – Open Space Valuation Panel
3:00-3:30	BREAK
3:30-4:45	W-133 Business Meeting
Wednesday	7:30-8:15 Continental Breakfast
8:15-9:45	Moderator - Tom Stevens
	Addressing Aggregation Bias in Zonal Recreation Models
	Klaus Moeltner. University of Nevada, Reno
	Single Bounded versus Multiple Bounded Estimators: Some Evidence on Efficiency Gains
	Daniel Hellerstein, USDA/ERS and Kevin Boyle, University of Maine
	Scope Sensitivity and the Validity of Contingent Valuation: A Preliminary Comparison of Aggregate and Individual Data
	Thomas A. Heberlein, University of Wisconsin—Madison; Matthew A. Wilson, University of Maryland; Richard C. Bishop and Nora C Schaeffer, University of Wisconsin—Madison
9:45-10:15	BREAK
10:15-11:45	Moderator - Jerry Fletcher
	Logit Models for Pooled Contingent Valuation and Contingent Rating/Ranking Survey Data- Valuing Benefits of Finnish Forest Biodiversity Conservation
	Juha Siikamki, University of California, Davis
	Estimating the Economic Value of Big Game Habitat Production from Natural and Prescribed Fire
	John Loomis, Dana Griffin and Ellen Wu, Colorado State University; Armando Gonzalez-Caban, Forest Fire Lab, Pacific Southwest Station, USDA Forest Service; and Daniel McCollum, Rocky Mountain Research Station
	Measuring the Value of Outdoor Recreation with Revealed and Stated Preference Data
	Jason Kinnell, Triangle Economic Research
12:00	ADJOURN

Too Many NOs? An Experimental Investigation of Behavior in Double Referenda

A.C. Burton,
Nuffield Institute for Health
University of Leeds

Katherine S. Carson,
United States Air Force Academy

Susan M. Chilton,
University of Newcastle

and W.G. Hutchinson
Queen's University, Belfast

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

This research is funded by the National Science Foundation (Division of Social and Economic Sciences, Decision, Risk, and Management Sciences Program, #SES-9986407). The opinions expressed herein are solely those of the authors and do not necessarily reflect the views of the National Science Foundation or the authors' institutions. Errors are the sole responsibility of the authors.

Abstract: Econometric evidence, theoretical speculations and the NOAA Panel on Contingent Valuation suggest that experimental evidence on potential biases in response to double referendum contingent valuation surveys is needed. This suggestion provides the starting point for this paper. There are several well-known stylized facts about responses to double-bounded referendum surveys. Carson et al. (1999, p. 22) summarize these stylized facts. They provide several theories of subject behavior that explain these stylized facts. We report the results of induced preference experiments designed to test two of these theoretical explanations versus a baseline of certainty and truthful preference revelation. The theories tested are simple cost uncertainty and weighted averaging. In the simple cost uncertainty case, the theory predicts more "No" responses to the second valuation question than expected under certainty and truthful preference revelation. In the weighted average case, the theory predicts more "Yes-Yes" and "No-No" responses to the two questions than expected under certainty and truthful preference revelation. The results find evidence to support the weighted average hypothesis. However, we find no evidence to support the uncertain cost hypothesis. This result merits further investigation.

This paper demonstrates how accepted methods from mainstream experimental economics can be utilized to test previously unverified explanations of econometrically observed behavior in double referendum contingent valuation surveys. If subject behavior matches the theoretical predictions, and if the same outcomes are observed in the field, then these experimental results provide information that may be of use in developing appropriate econometric techniques to apply to data from double bounded referendum surveys. In addition, these results may be helpful in redesigning double referendum surveys to avoid the respondent uncertainty and resulting bias typically associated with the follow up bid.

Introduction

In 1993, the NOAA Panel on Contingent Valuation recommended the use of the hypothetical referendum question format in surveys designed to estimate losses associated with natural resource damage. A hypothetical referendum question asks, for example, "If your cost for program A were \$D per year, would you vote for program A?" The panel recommended the hypothetical referendum format, because,

> ...as far as strategic reasons go, a respondent who would not be willing to pay D dollars has no reason to answer "Yes," and a respondent who would be willing to pay D dollars has no reason to answer "No." (Arrow et al., 1993, p. 4606).

Practitioners of contingent valuation typically employ a double-bounded, or dichotomous choice, referendum questions to generate estimates for willingness to pay. This referendum requires the respondent to answer two 'YES-NO' questions of the type: "Would you pay \$D for A?" The first question poses a value of \$D previously selected from a range of values. The second question poses a value taken from a sub-range (either higher or lower than \$D) that depends on the respondent's answer to the first question. Arguments favoring the use of the double referendum format in surveys center on its simplicity and the substantial gains in statistical efficiency compared to asking only one question (Hanemann, Loomis, and Kanninen (1991)).

One interesting stylized fact regarding double referendum surveys is an empirically observed systematic inconsistency of the responses to the first and second questions. Respondents' implicit point valuations of the resource at the moment of the first question often appear to differ from their valuations when the second question is asked. Furthermore, the error terms of the two valuations are not perfectly correlated and often have a correlation coefficient of less than unity. Therefore, the number of negative responses to the second question is often higher than would be expected from the distribution of values based on responses to the first question alone (Hanemann et al. (1991), McFadden and Leonard (1993)). A number of econometric methods to address this issue have been proposed (e.g. Cameron and Quiggin (1994), Alberini (1995)). Carson, Groves, and Machina (1999, hereafter CGM) also provide several theoretical explanations for these results.

Part of the Blue Ribbon Panel's recommendations about contingent valuation included an admonishment about the use of survey formats that deviate from the simple single referendum format. In its *Guidelines for Value Elicitation Surveys,* the panel stated, "If a double-bounded dichotomous choice or some other question form is used in order to obtain more information per respondent, experiments should be developed to investigate biases that may be introduced " (Arrow et al. (1993), p. 4612). Given the stylized facts described above and the concerns of the NOAA Panel regarding double-bounded dichotomous choice surveys, it is interesting that there has been no experimental research on the topic to date. Part of the reason for this is that there is still considerable debate surrounding the incentive compatibility of hypothetical single referendum surveys. The experimental evidence regarding the incentive compatibility of these surveys is decidedly mixed. Burton et al. (2001) summarize the experimental results in this area. Perhaps is it the view of the experimental community that the debate regarding single bounded referendum questions should be satisfactorily resolved before moving on to the more complicated, double bounded case.

Burton et al. (2001) report the results of a set of experiments using induced values to test the incentive compatibility of hypothetical single referendum survey questions. These experiments reveal little of the bias in responses to hypothetical and real question formats found in previous homegrown value-based experiments (e.g. Cummings et al. (1997)). Given that it appears that responses to single referendum contingent valuation surveys may indeed be accurate reflections of subjects' underlying values, it is now time to address the issues surrounding more complicated survey formats such as double bounded dichotomous choice. In this paper, we report the results of a set of experiments designed to test two of the four theoretical explanations proposed by CGM to explain the stylized facts about double bounded referendum surveys. We term these two theoretical explanations the Weighted Average hypothesis and the Uncertain Cost hypothesis. The purpose of these experiments is to determine if subject behavior is consistent with the theoretical predictions when the incentives described in the theories are induced. The results do provide support for the Weighted Average hypothesis. However, the results do not provide clear evidence regarding the Uncertain Cost hypothesis. These results provide insight into subject behavior when responding to double bounded contingent valuation surveys. They may generate new theoretical explanations of subject behavior as well as aid in the development of new empirical methods to adjust willingness to pay estimates from actual survey data.

Section II contains a description of the theories posed by CGM to explain observed responses to double bounded dichotomous choice surveys. The experimental design is described in Section III. We report the results in Section IV and present some concluding thoughts in Section V.

Theoretical Background

CGM hypothesize about possible subject behaviors that may explain the stylized facts surrounding double-bounded dichotomous choice contingent valuation questions. It is important to note that when willingness to pay distributions are estimated, it is assumed that subjects are not responding strategically to the valuation questions. In addition, it is assumed that subjects accept the costs presented to them as certain. Therefore, the estimated distributions of willingness to pay are generated under the assumptions of certainty and truthful preference revelation. A violation of either of these assumptions could generate the empirical regularities observed in contingent valuation data. All of the hypotheses in CGM address the issue of subject uncertainty. This uncertainty stems from the fact that at the start of the interview, subjects do not know that they will be asked more than one valuation question. The introduction of this additional question introduces uncertainty in the mind of the subject. The nature of subject uncertainty determines the optimal response. CGM hypothesize that there are four possible effects of the second valuation question, which we will term Uncertain Cost, Weighted Average, Bargaining, and Quantity/Quality Shift hypotheses. These four hypotheses are described in turn below.

In the Uncertain Cost hypothesis, subjects respond to the uncertainty created by the second valuation question by treating the second question independently from the first, but treating the new cost as an uncertain value with a mean equal to the second stated cost amount. Thus, respondents' risk preferences play a role in determining their answers to the second question. CGM assume that subjects are risk adverse. As a result, the distribution of stated willingness to pay will be skewed to the left as more subjects respond "No" to the second question than if the cost were known with certainty.

In the Weighted Average hypothesis, subjects respond to the second cost amount by assuming that the true cost is 'somewhere in the middle' of the two costs suggested by the survey administrator. As a result, subjects answer the follow up question as if the cost is a weighted average of the first and second costs, where the weights sum to one. This behavior results in responses to the follow up question that are more extreme than responses to a certain second cost. To see why, consider subjects whose values for the good lie in between the first and second costs. Under certainty, such a subject will respond either "No-Yes" or "Yes-No" to the two valuation questions, depending on whether the first offered cost is higher or lower than the subject's value. However, in the weighted average case, subjects who respond "No" initially are more likely to respond "No" a second time, since the weighted average of the two costs exceeds the lower bound of the cost range. Similarly, subjects who respond "Yes" initially are more likely to respond "Yes" to the follow up question as well since the weighted average is less than the upper bound of the cost range. This hypothesis predicts that there will be more "Yes-Yes" and "No-No" responses to the two questions than if the second cost were known with certainty.

The Uncertain Cost and Weighted Average hypotheses are observationally distinct and, as theories, are relatively clear. We test these two hypotheses in this paper. This is not true of the two further hypotheses offered in CGM, which we label the Bargaining hypothesis and the Quantity/Quality Shift hypothesis.

The Bargaining Hypothesis is based on the intuition that respondents may react to the second stated cost amount by assuming that they are now in a bargaining situation. This is particularly plausible if one remembers that the person implementing the survey is responding as if he or she were bargaining with the respondent (i.e. if the respondent says "Yes"' to the first stated cost amount, the cost amount is raised, but the cost amount is lowered if the respondent says "No"). CGM suggest this intuition implies that respondents are more likely to say "No" to the second cost amount because they hope it will lead to a lowering of any subsequent offers.

CGM do not offer an explicit bargaining model and their stated arguments imply the same predictions as the Uncertain Cost hypothesis. In fact, psychological intuitions aside, there appears to be no difference between these two hypotheses. Consider a two person bargaining situation where neither knows the other's preferences. When considering a response to any offer, the information available to a respondent is the set of possible cost options, a set of beliefs about those options (i.e. a probability distribution over the cost options) and preferences that include attitudes to the good and to risk. In other words, the model becomes similar to the model in the Uncertain Cost hypothesis. Thus, our experimental test of the Uncertain Cost hypothesis is also an indirect test of the Bargaining hypothesis, where it is assumed that the subject believes the cost amounts are normally distributed.

The Quantity/Quality Shift hypothesis is based on the intuition that respondents take price signals to be an indicator of the amount of good on offer or the quality of the good. When offered the second cost amount, the respondent takes this to indicate that a new good is being offered since it has a different price. Thus those who are offered a higher cost amount in the second round are more likely to say "Yes" because they feel that a new, better or bigger, good is being offered. Likewise, those who are offered a lower second cost amount are more likely to say "No" because they feel they are being offered a new, lower quality, or smaller, good.

We find it difficult to interpret this hypothesis from an economic perspective, although as stated, it appears observationally equivalent to the Weighted Average hypothesis. Consider respondents who are offered lower cost amounts, so that they believe that the good now being offered is of lower quality. *A priori,* their responses depend upon their preferences over this good relative to other goods, their beliefs about the quality (or quantity) change and the amount the price has changed. In fact, if the price has fallen enough, it may *increase* the likelihood of wanting to pay for the good. This is another way of saying that the response to a change in price and quantity of a good depends on the own-price elasticity of demand for that good. Without further information, we feel this hypothesis requires very strong assumptions about the price elasticity of environmental goods in order to make clear predictions about responses. Thus, we do not test this hypothesis in the experiment.

Table 1 summarizes the predictions of these four hypotheses. There are four possible response combinations to the two valuation questions: "Yes-Yes," "Yes-No," "No-Yes," and "No-No." The first column depicts the number of each type of response predicted under certainty and truthful preference revelation. A, B, C, and D represent non-negative amounts of each response type that depend on the distribution of values for the public good in the sample population. The values in columns 2, 3, 4, and 5 reflect how the observed responses for each hypothesis are predicted to differ from the certainty, truthful preference revelation baseline. Clearly, the Uncertain Cost and Bargaining hypotheses predict similar deviations from the baseline, as do the Weighted Average and Quantity/Quality Shift hypotheses. Columns 1, 2, and 4 are relevant to the description of the experimental design that follows. Because these four hypotheses result in only two distinct vote distributions, we test only two hypotheses: Uncertain Cost and Weighted Average.

Table 1. Double Referendum Response Effects

	Certainty, Truthful Preference Revelation (Baseline)	**Uncertain Cost**	**Bargaining**	**Weighted Averaging**	**Quantity/ Quality Shift**
Yes-Yes	A	< A	< A	> A	> A
Yes-No	B	> B	> B	< B	< B
No-Yes	C	< C	< C	< C	< C
No-No	D	> D	> D	> D	> D

Experimental Design

One of the difficulties with replicating a double-bounded referendum in an experimental laboratory is that subjects in the field do not know that they are going to be asked a second valuation question until the moment that the question is posed to them. This aspect of contingent valuation surveys generates problems in a controlled laboratory setting where, to conform to proper experimental protocol, subjects must be fully informed of experimental procedures from the outset. Telling subjects at the beginning that they will be asked two questions creates two problems. First, it can generate strategic responses to the first question, particularly if subjects know that they will receive a lower cost offer in the second question if they respond "No" to the first question. Since the baseline case in the hypotheses requires truthful preference revelation, this outcome is undesirable. Second, if subjects know that they will be asked a second question, it is more difficult to create the types of uncertainty posited in the hypotheses to be tested. This experimental design avoids these problems by employing a "virtual first round." Subjects play only one round under the (experimental) assumption that they have played a previous one-shot round in which they truthfully revealed their preferences, consistent with the (econometric) assumption that respondents truthfully reveal their preferences when responding to the first bid amount in a referendum contingent valuation study (Cameron and Quiggin (1994) employ such an assumption in their analysis).

Thus, the first round is virtual in the sense that subjects do not actually play it. However, given their assigned value and cost and the assumption of truthful preference revelation, it is possible to predict what subjects would have voted in this round. This "virtual vote" then determines the cost that subjects are offered in the second, or actual round, which is the only round that subjects play. For example, if a subject's assigned value exceeds the initial offered cost, his or her virtual vote is "Yes." The subject is then offered a cost higher than the initial cost in the actual round. If a subject's assigned value is less than the first offered cost, his or her virtual vote is "No." The subject will then be offered a cost lower than the initial cost in the actual round. In this way, the experimental design is consistent with the assumption of truthful preference revelation in response to the first valuation question employed in the literature on double-bounded dichotomous choice. As noted, Burton et al. (2001) provide experimental evidence to support this assumption.

The experimental design has four main features: treatments, experimental procedures and instruction packets, induced values, and sample sizes and subject recruitment. These four elements of the experimental design are described below.

Treatments

The experimental design consists of three treatments: the Baseline treatment, the Uncertain Cost treatment, and the Weighted Average treatment. In each treatment, the basic design of the experiment is the same. Each experimental session consists of 9 subjects per treatment for a total of 27 subjects per experimental session. Subjects in a session are randomly assigned to a treatment. In each treatment, each of the 9 subjects is a shareholder in an imaginary investment company called "Experimental Investments Limited (EIL)." The group must vote on whether EIL should make an investment for the group. Each shareholder begins the experimental session with a total of 10 tokens. Each shareholder knows how his or her cost of the investment will be determined and personal return (**R**) from the investment. Assigned values of costs and **R** are discussed in the section describing the induced values below. EIL makes the investment if at least 6 of the 9 shareholders vote "Yes." If less than 6 shareholders vote "Yes," then the session terminates and each shareholder completes the experiment with 10 tokens. If the vote passes, then shareholders' final token balance equals 10 minus their investment amount plus their return from the investment. At the completion of a session, tokens are converted to dollars at the rate of $1.00 per token. The only variable that differs between treatments is how shareholders' individual costs of the investment are determined.

In the Baseline treatment, shareholders know their cost of the investment with certainty. Therefore, shareholders whose **R** exceeds their assigned cost are expected to vote "Yes" and shareholders whose **R** is less than their assigned cost are expected to vote "No." This treatment serves as the control and corresponds to column 1 of Table 1.

In the Uncertain Cost treatment, shareholders know that their cost of the investment will be drawn from a normal distribution. Although CGM do not explicitly state that respondents' uncertainty about the cost takes the form of a normal distribution, we take as our starting point the view that any theory that formalizes this uncertainty would be unlikely to exclude the normal distribution. In the experiment, the normally distributed cost has a mean of **M** and a standard deviation of 5. Shareholders know their assigned value of **M** and see a table illustrating how their cost of the investment will be determined based on a drawing from a bag of cards numbered from 1 to 100. This table is reproduced as Table 2. The drawing occurs after the vote and only if 6 of the 9 shareholders vote "Yes." The mean value, **M**, equals the assigned cost that is known with certainty in the Baseline treatment. Therefore, this treatment replicates the incentive structure described in the Uncertain Cost hypothesis and corresponds to column 2 of Table 1.

In the Weighted Average treatment, shareholders know that their cost of the investment will be either a low value (**L**) or a high value (**H**). Shareholders know their values of **L**, **H** and **R**. They also know that their cost will be determined by two drawings from a bag of cards numbered from 1 to 100. The first drawing occurs before the vote and determines the *threshold number*. The *threshold number* indicates whether it is more likely that the shareholders' investment cost will be **L** or **H**. If the vote passes, a second drawing from the bag determines whether shareholders' cost is **L** or **H**. If a number from one to the *threshold number* is drawn, then the investment cost is **L**. If a number from the *threshold number plus one* to 100 is drawn, then the investment cost is **H**. Shareholders see a table summarizing the rules of the drawing, which is included as Table 3. The values of **L** and **H** correspond to offered costs in first and second valuation questions. It is important to note that for some shareholders, **L** equals the cost from the virtual round, while for other shareholders **H** equals the cost from the virtual round. Whether a shareholder's virtual cost is **L** or **H** depends on his or her assigned value of **R**. This treatment replicates the incentive structure described in the Weighted Average hypothesis and corresponds to column 4 of Table 1.

Table 2. Distribution of Possible Investment Amounts (C) – Uncertain Cost Treatment

If the card number is:	**Then the value of C is:**
1	**M -10**
2 or 3	**M - 9**
4 or 5	**M - 8**
6, 7, or 8	**M – 7**
9, 10, 11, or 12	**M - 6**
13, 14, 15, 16 or 17	**M - 5**
18, 19, 20, 21, 22, or 23	**M - 4**
24, 25, 26, 27, 28, 29, or 30	**M - 3**
31, 32, 33, 34, 35, 36, 37, or 38	**M - 2**
39, 40, 41, 42, 43, 44, 45, or 46	**M - 1**
47, 48, 49, 50, 51, 52, 53, or 54	**M**
55, 56, 57, 58, 59, 60, 61, or 62	**M + 1**
63, 64, 65, 66, 67, 68, 69, or 70	**M + 2**
71, 72, 73, 74, 75, 76, or 77	**M + 3**
78, 79, 80, 81, 82, or 83	**M + 4**
84, 85, 86, 87, or 88	**M + 5**
89, 90, 91,or 92	**M + 6**
93, 94, or 95	**M + 7**
96 or 97	**M + 8**
98 or 99	**M + 9**
100	**M + 10**

Table 3. Distribution of Possible Investment Amounts – Weighted Average Treatment

If the card that we draw in the second drawing is numbered	Then your investment amount is:
From 1 to the *threshold number.*	**L**
From the *threshold number plus one* to 100	**H**

Experimental Procedures and Instruction Packets

At each experimental session, 27 students arrive at that designated room at the designated time. Security measures are in place to ensure that no subject participates more than once. As subjects enter the room, they draw a colored card from an envelope containing cards of three different colors. The color of the card determines in which treatment the subject participates. Once everyone has a card, each group of 9 follows the designated moderator for that group to a different room and the experiment begins. By running all three treatments concurrently at each experimental session, we can reduce session-specific effects by randomly distributing subjects among treatments.

Once in the different rooms, each subject receives a blank envelope containing an instruction packet and a smaller envelope, which contains a ballot slip and "Investment Return Slip." The instructions for each treatment differ in the description of how shareholders' investment costs are determined. The relevant portion of the Baseline treatment instructions state:

> Each of you is a shareholder in Experimental Investments Limited (EIL). As a shareholder, you will have the opportunity to vote on whether EIL will invest a number of tokens on your behalf. To help you understand the investment we will use the following abbreviations:
>
> **C** = your investment amount in tokens
>
> **R** = your return from the investment in tokens
>
> If EIL votes to make the investment, then **C** tokens will be deducted from your account balance. You will know your value of **C** before you vote on the investment decision. Different people in the room have different values of **C**. This value of **C** is private information. Do not reveal your value of **C** to the other people in the room.
>
> If EIL votes to make the investment, then **R** tokens will be added to your account balance.
>
> You will know your value of **R** before you make the investment decision. Different people in the room have different values of **R.** This value of **R** is also private information. Do not reveal your value of **R** to the other people in the room.
>
> To summarize, if EIL votes to make the investment then **C** tokens will be deducted from your account balance and **R** tokens will be added to your account balance. When you make your investment decision you will know your values of **C** and **R**.

By contrast the Uncertain Cost treatment instructions state:

Each of you is a shareholder in Experimental Investments Limited (EIL). As a shareholder, you will have the opportunity to vote on whether EIL will invest a number of tokens on your behalf. To help you understand the investment we will use the following abbreviations:

- **C** = your investment amount in tokens
- **M** = the average value of **C** for you
- **R** = your return from the investment in tokens

If EIL votes to make the investment, then **C** tokens will be deducted from your account balance. The value of **C** depends upon your value of **M** and the results of a drawing that I will conduct if EIL votes to make the investment. I will explain how the drawing will work on the next page.

You will know your value of **M** before you vote on the investment decision. Different people in the room have different values of **M**. This value of **M** is private information. Do not reveal your value of **M** to the other people in the room.

If EIL votes to make the investment, then **R** tokens will be added to your account balance. You will know your value of **R** before you vote on the investment decision. Different people in the room have different values of **R**. This value of **R** is also private information. Do not reveal your value of **R** to the other people in the room.

The Drawing

The outcome of a drawing will determine your investment amount. We will only conduct this drawing if EIL votes to make the investment decision.

The bag that I am holding contains 100 cards numbered from 1 to 100. You may verify the numbers on the cards at any time during the experiment if you so desire. If EIL votes to make the investment, I will ask one of you to draw one card from this bag and announce the number on the card. This number will determine your investment amount, **C**.

The table below explains how the value of **C**, your investment amount, depends upon the outcome of the drawing. For example, if the number on the card drawn is 33, then your investment amount is **M-2**. If the number on the card drawn is 86, then your investment amount is **M + 5**.

You might find it helpful to know that the numbers in this table are based on a normal distribution with a mean of **M** and a standard deviation of 5.

The table described above is reproduced as Table 2. The relevant portion of the Weighted Average instructions state:

Each of you is a shareholder in Experimental Investments Limited (EIL). As a shareholder, you will have the opportunity to vote on whether EIL will invest a number of tokens on your behalf. To help you understand the investment we will use the following abbreviations:

- **L** = your low investment amount in tokens
- **H** = your high investment amount in tokens
- **R** = your return from the investment in tokens

If EIL votes to make the investment, then either **L** or **H** tokens will be deducted from your account balance. The outcomes of two drawings will determine whether the investment amount deducted from your account is **L** or **H**. I will explain how the drawings will work on the next page.

You will know your values of **L** and **H** before you vote on the investment decision. Different people in the room have different values of **L** and **H**. These values of **L** and **H** are private information. Do not reveal your values of **L** and **H** to the other people in the room.

If EIL votes to make the investment, then **R** tokens will be added to your account balance. You will know your value of **R** before you vote on the investment decision. Different people in the room have different values of **R**. This value of **R** is also private information. Do not reveal your value of **R** to the other people in the room.

The Drawings

The outcomes of two drawings will determine whether your investment amount is **L** or **H**. The purpose of the first drawing is to determine the *threshold number*. The *threshold number* indicates whether it is more likely that your investment amount will be **L** or that your investment amount will be **H**. We will conduct this drawing before you vote to make the investment decision.

The purpose of the second drawing is to determine which amount, **L** or **H**, is your investment amount. We will only conduct the second drawing if EIL votes to make the investment decision.

The bag that I am holding contains 100 cards numbered from 1 to 100. You may verify the numbers on the cards at any time during the experiment if you so desire. Before EIL votes on the investment decision, I will ask one of you to draw one card from the bag and announce the number on the card. This number will determine the *threshold number*.

The table below explains how the *threshold number* indicates the likelihood that your investment amount will be **L** or **H**. If EIL votes to make the investment, we will conduct a second drawing to determine whether your investment amount is **L** or **H** in the same manner as we conduct the first drawing.

If the card that we draw in the second drawing is numbered from 1 to the *threshold number*, then **L** tokens will be deducted from your account balance.

If the card that we draw in the second drawing is numbered from the *threshold number plus one* to 100, then **H** tokens will be deducted from your account balance.

You might find it helpful to know that the likelihood that your investment amount is **L** or **H** is based on a weighted average of **L** and **H**.

The table described above is reproduced as Table 3. All versions of the instructions continue with examples of how to compute payoffs in tokens and quizzes that test the subjects' ability to apply the voting rule, use the results of the drawings to compute their investment amounts, compute their earnings in tokens, and convert these earnings to dollars. Subjects do not learn their individual values of **R** or **C**, **M**, or **L** and **H** (depending on treatment) until they open the envelope containing their ballot slip. Each subject's **R**, and **C**, **M**, or **L** and **H** are private information. Copies of the experimental instructions and supporting materials are available from the authors upon request.

Induced Values

There are three types of shareholders. Low value shareholders have an **R** of 16, high value shareholders have an **R** of 26, and very high value shareholders have an **R** of 32. Very low value shareholders, whose **R** would be less than both the first and second offered costs, are not included in the design because all hypotheses predict that they will vote "No" in the actual round. All shareholders' offered cost in the virtual round is 20. Therefore, the virtual vote of the low value shareholders in "No," while that of the high and very high shareholders is "Yes." In the Baseline treatment, low value shareholders have a **C** of 10. High and very high shareholders' **C** is 30. Therefore, we expect the low and very high shareholders to vote "Yes" in the Baseline treatment, and the high value shareholders to vote "No." The values of **M** for shareholders in the Uncertain Cost treatment are the same as the values of **C** for each shareholder type in the Baseline treatment. However, due to the fact that the investment amount is not known with certainty, we can expect more Low and Very High shareholder types to vote "No" in this treatment.

In the Weighted Average treatment, low value shareholders have an **L** of 10 and an **H** of 20. We expect some of these shareholders to vote "No." High and very high value shareholders have an **L** of 20 and an **H** of 30. We expect some of the high value shareholders to vote "Yes." Since **R** for the Very High value shareholders exceeds **H**, we expect them to vote "Yes" as well.

There are 72 subjects per treatment, 24 of each value type. We employ this distribution of values, in conjunction with the 2/3 voting rule for three reasons:

(1) If subjects behave in accordance with the predictions of the hypotheses, it is more likely that the votes in the Baseline and Weighted Average treatments will pass and the vote in the Uncertain Cost treatment will fail. This results in higher payoffs for most subjects and positive externalities from word of mouth in terms of recruiting future subjects.

(2) It makes it more difficult for subjects who don't entirely understand the rules to sway the results. If the voting rule were 50%, one confused subject could lower everyone's payoff, particularly in the Baseline or Weighted Average treatments.

(3) It provides more useful data for analysis. In both the Uncertain Cost and Weighted Average treatments, two of the three subject types are predicted to have vote distributions that deviate from the distribution of votes by subjects of the same type in the Baseline treatment. Therefore, we have can employ 2/3 of our sample for each hypothesis test.

Although one can ask subjects not to discuss the experiment with others after the experimental session, it is probably unreasonable to assume that no outside communication takes place. Therefore, the design contains several safeguards to make such communication irrelevant. First, although the distribution of value types is set at 1/3 of each type across the eight experimental sessions, the distribution of types in any given treatment in any given experimental session is determined by a draw from this distribution. In this way, the investment may be made on some days and not on others. Since subjects were permitted to view the ballot sheets to verify the moderator's count, the distribution of types was apparent to them. Some of the subjects did inquire about whether the distribution was always the same. The moderators replied that it changed every day and did not reveal the overall distribution of values from which the daily allocation was drawn.

The second safeguard that was employed was to change the color of the card that determined each treatment every day. For example, on one day a blue card might mean that a subject plays in the Uncertain Cost treatment and on another day it might mean a subject plays in the Baseline treatment. In this way, information like, "Try to get a blue card," would be useless to a potential subject.

Sample Sizes and Subject Recruitment

A power analysis was conducted using a range of assumptions about observed deviations between treatments, ranging from small to large. Under very conservative assumptions, a sample size of 200 results in a power of approximately 0.80 at a level of significance of 0.05. At this sample size, the power approaches one rapidly as deviations between treatments increase. Because each treatment requires 9 subjects and treatments are run concurrently, the sample size must be divisible by 27. Therefore, the experimental design requires a sample size of 216. All subjects were undergraduate students the United States Air Force Academy.

Subject recruitment was accomplished using posters and flyers and announcements in classes. Subjects were informed of the amount of time the experiment would take (approximately 30-45 minutes) and average earnings ($10.00).

Results and Discussion

Table 4 reports the predicted distributions of "Yes-Yes," "Yes-No," "No-Yes," and "No-No" for each treatment based on the induced values and costs and the predictions of the hypotheses proposed by CGM. The high and very high value subjects' values and costs determine the predicted numbers of "Yes-Yes" and "Yes-No" votes. The low value subjects' values and costs determine the predicted numbers of "No-Yes" and "No-No" votes. Note that as described above, there are no predicted "No-No" votes in the Baseline treatment because we omit this subject type from the experimental design. However, both the Uncertain Cost and Weighted Average hypotheses predict a positive number of "No-No" votes in this treatment. Therefore, nonzero levels of "No-No" votes are one form of evidence for behavior that supports both hypotheses.

Table 5 reports the observed vote distributions in each treatment. A cursory examination of the table indicates that subjects do behave in accordance with the predictions of economic theory in the case of the Baseline treatment. Furthermore, the vote distribution in the Weighted Average treatment does move in the direction predicted by the hypothesis. However, this is not the case for the Uncertain Cost treatment. In particular, there are zero "No-No" votes in this treatment. While this hypothesis predicts that the distribution will be shifted to the left of that resulting from the Baseline treatment, the data seem to indicate a shift to the right. We can quantify these observations by testing the following hypotheses using nonparametric tests:
Hypothesis 1: The vote distributions in the Baseline and Uncertain Cost treatments are the same
Hypothesis 2: The vote distributions in the Baseline and Weighted Average treatments are the same Table 6 reports the results of these tests.

We find no evidence to support Hypothesis 1. In fact, we can only reject this hypothesis at the 10% level of significance ($p = 0.072$). There are two explanations for the lack of evidence for this hypothesis. The first is that subjects simply do not behave in accordance with the theoretical predictions. The second is that one of the assumptions of the theory is violated. In particular, CGM argue that one will only observe an increase in "No" responses to the second question for risk adverse subjects. Therefore, it is possible that this subject pool contains few risk adverse subjects. This possibility can be verified using a very simple test. A useful exercise would be to repeat this experiment on a group of subjects who have been determined to be risk adverse using this test to determine if their behavior then matches the predictions of the theory.

In summary, there appears evidence to suggest that subjects do not invoke the Uncertain Cost hypothesis. This hypothesis is further weakened as a useful explanation of general respondent reaction to the second cost amount by the fact that the observed distribution of responses is in the opposite direction to that proposed by CGM.

We find strong evidence to support Hypothesis 2. We can reject the null hypothesis that the vote distributions in the Baseline and Weighted Average Treatments are the same at any standard level of significance (p = 0.003). We further note that the distributions is as predicted by CGM. Therefore, subjects do behave in accordance with the predictions of the Weighted Average hypothesis. If subjects take a weighted average of the presented costs when responding to a field survey, then one should expect to obtain estimated willingness to pay distributions that are consistent with our results. Furthermore, econometric methods that incorporate this information about subject behavior into the estimation procedure should result in more accurate willingness to pay estimates.

Table 4. Predicted Responses in Double Referendum Treatments

	Baseline	**Uncertain Cost**	**Weighted Average**
Vote Type			
Yes-Yes	24	< 24	> 24
Yes-No	24	> 24	< 24
No–Yes	24	< 24	< 24
No-No	0	> 0	> 0

Table 5. Observed Responses in Double Referendum Treatments

	Baseline	**Uncertain Cost**	**Weighted Average**
Vote Type			
Yes-Yes	24	35	36
Yes-No	24	13	12
No–Yes	24	24	18
No-No	0	0	6

Table 6. Nonparametric Test Results

Hypothesis 1 – H_0: Baseline = Uncertain Cost	
Pearson's Chi-Squared: p = 0.070	Fisher's Exact: p = 0.072
Hypothesis 2 – H_0: Baseline = Weighted Average	
Pearson's Chi-Squared: p = 0.004	Fisher's Exact: p = 0.003

Conclusions

The experiment reported above was not designed to prove conclusively which, if any, of the tested hypotheses are adopted by individual respondents when faced with a follow up bid in a double referendum. What we have done is to test the two proposed hypotheses under strict laboratory conditions to examine whether, when strongly induced, they are capable of generating the types of voting behavior observed in the contingent valuation literature. Our experimental findings, while not testing every possible explanation of apparently inconsistent observed responses, are a useful starting point in providing much needed empirical evidence in the area.

The implications of the results from the Weighted Average treatment are straightforward. If such behavior is present within a particular sample of contingent valuation survey respondents then there will be significant differences in voting behavior over what would be the case compared to certain costs (i.e. our Baseline treatment). This difference is in the direction proposed by CGM. It is, however, difficult to draw definitive conclusions with respect to the Uncertain Cost hypothesis. As currently formulated CGM's theory predicts that for risk adverse subjects the difference in voting behavior between the Baseline and Uncertain Cost treatments will result in more "No" responses to the second question, as illustrated in Table 1. Although we did observe significant differences between treatments, they were in the opposite direction to those proposed by CGM. A sample that included a sufficient number of risk loving respondents would produce such a result. To date, as far as we are aware, there is no evidence regarding the risk preferences of past contingent valuation survey respondents. Therefore, it is difficult to say whether the Uncertain Cost hypothesis has had any impact on contingent valuation survey results, whether in the direction predicted by CGM or in the direction found in our study.

In closing we note that this paper demonstrates how appropriate procedures from mainstream experimental economics can be utilized to test previously unverified explanations of econometrically observed inconsistent voting in double referenda contingent valuation surveys. If subject behavior matches the theoretical predictions, and if the same outcomes are observed in the field, then these experimental results provide information for use in developing appropriate econometric techniques to apply to data from double bounded referendum surveys. In addition, these results may be helpful in redesigning double referendum surveys to avoid the respondent uncertainty and resulting bias typically associated with the follow up bid.

References

Alberini, Anna. "Efficiency vs. Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models." *Journal of Environmental Economics and Management* 29, no. 2 (September 1995): 169-80.

Arrow, Kenneth J.; Solow, Robert; Portney, Paul; Leamer, Edward E.; Radner, Roy; and Schuman, Howard. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58, no. 10 (January 15, 1993): 4602-14.

Burton, A.C.; Carson, Katherine S.; Chilton, Susan M.; and Hutchinson, W.G. "Are Inconsequential Referenda Incentive Compatible?" Paper presented at the Allied Social Science Association Annual Meetings, January 5-7, 2001, New Orleans, LA.

Cameron, Trudy Ann and Quiggin, Jon. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-up' Questionnaire." *Journal of Environmental Economics and Management* 27, no. 3 (November 1994): 218-34.

Carson, Richard; Groves, Theodore; and Machina, Mark. "Incentive and Informational Properties of Preference Questions." Plenary Address to the European Association of Resource and Environmental Economists, Oslo, Norway, June 1999.

Cummings, Ronald G.; Elliott, Steven; Harrison, Glenn W.; and Murphy, James. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* 105, no. 3 (June 1997): 609-621.

Hanemann, Michael; Loomis, John; and Kanninen, Barbara. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *American Journal of Agricultural Economics* 73, no. 4 (November 1991): 1255-63.

McFadden, Daniel and Leonard, Gregory K. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." in Hausman, Jerry A., ed. *Contingent Valuation: A Critical Assessment.* Amsterdam: North-Holland, 1993.

Testing the Validity of Contingent Behavior Trip Responses

Therese Cavlovic
Assistant Professor
Department of Economics
Weber State University

Robert P. Berrens
Associate Professor
Department of Economics
University of New Mexico

Alok K. Bohara
Professor
Department of Economics
University of New Mexico

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

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Address Correspondence to T. Cavlovic, Economics Department, Weber State University, 3807 University Circle, Oregon, UT 84408-3807; Tel: (801)626-6066.

Abstract: Following the prompting of Arrow et. al. (1993) and others, the number of validity tests of contingent valuation data has grown rapidly. However, to date, only several studies have examined the validity of contingent behavior data. The objective of this study is to take advantage of a unique opportunity to test the validity of contingent behavior trip data on rock climbing trips to Hueco Tanks, a premier rock climbing destination. A construct validity test of scope is conducted using data from surveys implemented before and after a policy restricting recreational access was imposed. Results from a generalized Negative Binomial regression model suggest that contingent behavior data may be a valuable supplement to revealed preference data when policy proposals are outside the range of historical conditions.

Key Words: Contingent Behavior, Rock Climbing, Test of Scope

Introduction

A recent trend in recreation demand modeling is to use contingent behavior (CB) trip data to value changes in consumer welfare under hypothetical scenarios, such as changes in management rules or environmental quality. Commonly, CB data is also combined with revealed preference (RP) data on past use levels. By definition, applications of CB questions are restricted to consideration of hypothetical use levels, and thus the measurement of use values. While potentially avoiding some of the criticisms (e.g., lack of familiarity with the good) concerning the application of contingent valuation (CV) methods and the measurement of non-use values, CB data still remains controversial due to its inherent hypothetical nature. However, given the restricted focus on use values, patterns of evidence concerning the validity of CV may not hold for CB data. Further, while CB applications have grown, there are few tests of CB validity. Thus, there is a considerable opportunity for insights from targeted CB validity studies, such as tests of scope, and comparisons of hypothetical and real behavior.

This study takes advantage of a unique opportunity to test the validity of CB data for outdoor rock climbing demand. Hueco Tanks Texas State Park, located outside of El Paso Texas, is known throughout the world as a premier climbing destination. In 1998, Texas Parks and Wildlife Department (TPWD) severely restricted open-recreational access at Hueco Tanks. TPWD believed that increased popularity of Hueco Tanks as a unique climbing destination threatened the park's ecological and cultural resources. For alternative access restrictions, a construct validity test of scope is conducted using data from surveys implemented both before and after the policy change.

The first survey was conducted in the spring of 1998, with the follow-up in the spring of 1999. In the 1998 survey, climbers who had visited Hueco Tanks were surveyed about their actual rock climbing trips <u>and</u> intended trips under alternative hypothetical policy rules restricting access (i.e., CB trip data). The 1999 survey was administered after access restrictions were imposed; climbers were surveyed about their actual post-policy rock climbing trips.

A construct validity test of scope is conducted comparing post-policy revealed preference (RP) trip data obtained from the 1999 survey and pre-policy RP and CB data obtained from the 1998 survey. To do a test of scope, each trip response is based on different levels of site access: pre-policy RP trip data are based on the least restrictive access policy; pre-policy CB trip data are based on gradual restrictions in site access; and post-policy RP data are based on the most restrictive access policy. Results from a pooled generalized Negative Binomial regression model suggest that CB data may be a valuable supplement to RP data when policy questions are outside the range of historical conditions. The value of access is significantly sensitive to scope.

Climbing at Hueco Tanks

While rock climbing has existed on public lands for the past century, recreational demand for climbing is perceived to have grown significantly over the last several decades. This growth has lead to a variety of new climbing management and access proposals (NPS 1993). Severe restrictions in access can cause significant loss in economic value to rock climbers. Hueco Tanks State park in Texas is a prominent example.

During the 1980s and 1990s, Hueco Tanks became known to climbers living throughout the world as a premier climbing destination providing numerous types of climbs, and what are referred to as boulder problems. Hueco Tanks is particularly famous for its quality and quantity of boulder problems, and ideal winter climbing conditions (i.e., dry and warm).

Unlike most types of rock climbing, bouldering does not require ropes, climbing protective gear, or knowledge about climbing protection. Strong, agile climbers climb on boulder problems generally not higher than 25 feet. Foam crash pads (approximately three inches thick and nine square feet) and spotters (i.e., other climbers) protect climbers from a fall. Climbers can generally walk off the back of boulders to descend. The "V" grading system is used to identify the difficulty of boulder problems and a climber's ability level (e.g., the ratings range from V0 through V14, where V0 represents the easiest rated boulder problem).

Due to increases in recreational use (primarily rock climbing) during the 1980s and 1990s, TPWD became concerned about the recreational impacts on park resources: Park planners with TPWD began to realize, even as they planned for increased recreational use…, that conflicts were going to occur between park users and there was a great need to protect the priceless rock art found throughout the park. The place was literally being loved to death by thousands of hikers, climbers, and picnickers. Increasing use by rock climbers from around the world is beginning to impact the park permanently…(Hueco Tanks State Historic Park 1997). In 1997, TPWD proposed a management plan recommending gradual restrictions in open-recreational access (TPWD 1997). On September 1, 1998, TPWD closed three of four mountain areas in Hueco Tanks to open-recreational access (TPWD 1998). Consequently, TPWD has greatly reduced access to a unique, world-class bouldering area.

Nonmarket Valuation and Contingent Behavior

Nonmarket valuation of environmental goods and services can be divided into revealed preference (RP) and stated preference (SP) approaches. RP approaches, such as the travel cost method (TCM), rely on observed individual behavior, often *revealed* in survey instruments, to infer values for environmental goods or services. A variety of stated preferences (SP) techniques are used to assess the economic value of nonmarket environmental goods. These methods include contingent valuation (CV) and contingent behavior (CB). In CV, respondents are asked to make statements about their willingness-to-pay (WTP), or to accept compensation, for changes in environmental quality.

CB is commonly used to assess quality or price changes at a recreational site. In the CB framework, respondents are asked to make statements about their intended behavior (e.g., visitation to a site) given a proposed change (e.g., in site quality, access, or price). Whereas CV elicits a value statement, CB is used to estimate changes in behavior or levels of use for a nonmarket good. For example, as part of the draft Environmental Impact Statement for the potential removal of the four Lower Snake River dams, Loomis (1999) uses CB trip data to estimate recreational benefits of this river restoration project to anglers and non-anglers.

A recent trend in recreation demand modeling is to combine RP and SP (RP-SP) trip data (Englin and Cameron 1996; Eiswerth et al. 2000; Loomis 1999; and Rosenberger and Loomis 1999). In the combined RP-SP recreation demand framework, individuals are asked to provide information on actual trips taken to a site under existing resource conditions or management rules (i.e., RP data), and subsequently asked to indicate the number of trips they would take to the site under alternative, hypothetical management rules (i.e., CB data). Similar to CV methods, use of CB data is controversial. Critics question the validity of SP techniques by arguing that respondents cannot accurately identify true statements about hypothetical WTP (Vatn and Bromley 1995) or intended visitation (Cicchetti and Peck 1989).

Validity is commonly thought of as accuracy in measurement (Loomis 1993). Construct validity involves the degree by which a measure relates to other measures as predicted by theory

(American Psychological Association 1974; Mitchell and Carson 1989, 191). For example, suppose an individual is confronted with a change in the level or scope of an environmental good from Q^0 to Q^1, where $Q^0 > Q^1$. Given strictly positive marginal utility for the good, then it is expected that the individual would value Q^0 more than Q^1 (Carson and Mitchell 1995, 156).

Following the prompting of Arrow et al. (1993) and others, the number of validity tests of CV has grown rapidly. These include considerable numbers of criterion (or external) validity tests (e.g., see review in Vossler and Kerkvliet 1999), and construct validity tests, such as tests of scope or temporal reliability (e.g., see reviews in Carson 1997; Carson et al. 1999).

While a growing number of researchers have examined the validity of CV methods, careful CB validity tests remain rare (Eiswerth et al. 2000; Loomis 1993; Nestor 1998). To date, the limited evidence provides qualified support for the use of CB questions and approaches.

Ideally, to assess the criterion validity of CB data that is used to estimate recreational benefits, researchers would like to compare CB trip responses that map into observed trips given an identical policy change (Berrens and Adams 1998; Loomis 1993). Alternatively, post-policy RP data could be compared to pre-policy CB responses. Difficulties arise in making such comparisons. For example, while *ex post* visitation data may exist, the actual change in site conditions may differ from the exact policy change proposed in CB questions. Further, the period between CB trip responses and *ex post* trips could vary considerably, in which individual preferences could have changed.

The Survey Method and Validity Test

An intercept plus follow-up mail survey was conducted in 1998 to collect data from climbers about their (pre-policy change) rock climbing trips and intentions to visit Hueco Tanks under alternative rules restricting access; the data account for over 2000 RP trips. The survey also included questions regarding details of climbers' trips to Hueco Tanks, including length of stay, lodging and travel expenses, travel accommodations (e.g., by car or airplane), the number of people traveling together on a trip, climber preferences for different climbing areas in Hueco Tanks, and purposes of visiting Hueco Tanks.

The survey was mailed first class to 752 climbers. A follow-up reminder letter was sent to nonrespondents four weeks after the original survey mailing. In addition, a follow-up survey and reminder letter was mailed to 100 random climbers who had not yet responded. The adjusted response rate (adjusted for undeliverable surveys) was 56 percent.

For pre-policy RP (PRE-RP) trip data, unrestricted access at Hueco Tanks included the following conditions: (1) all four mountain areas in the park were open to recreational access; (2) climbers were not allowed to climb in pictograph areas; (3) the park was limited to 60 vehicles, but made no restrictions on the number of individuals in the park at any one time; and (4) the entrance fee was $2, yet climbers had to pay an additional $2 activity fee (fees were reduced if climbers purchased an annual Texas Conservancy Passport).

The survey also included pre-policy CB questions. The CB questions read as follows:

> "Given West Mountain only is closed...Would your trips next season change because of this new policy? [If yes] You stated your trips would change. About how many more or fewer trips would you take next season?"

And,

> "Next suppose TPWD eliminates climbing access to both East and West Mountain...Would your trips next season change because of this new policy? [If

yes] You stated your trips would change. About how many more or fewer trips would you take next season?"

Throughout this paper, responses to the first and second set of CB questions will be referred to as SP1 and SP2 trip data, respectively. All climbers who had participated in the 1998 survey returned their survey prior to September 1, 1998.

When TPWD restricted open-recreational access at Hueco Tanks, climbers could still visit the park, but their climbing opportunities were limited. The rule limited access in the following ways: (1) open-recreational access was limited to North Mountain non-pictograph areas only, yet to guarantee a visit North Mountain, visitors had to call in advance to make a reservation; (2) North Mountain was limited to 50 visitors; (3) the total entrance fee remained at $4 (the fee was reduced if climbers purchased an annual Texas Conservancy Passport); (4) before entering, all visitors had to attend a mandatory park orientation; and (5) visitors were allowed to be guided by a park ranger to the remaining three mountain areas—East Mountain, West Mountain, and East Spur Maze—for a two hour period.

The change in access provided a unique opportunity to question original survey respondents about their trips to Hueco Tanks under this new rule. The second survey, which was mailed to climbers one year after the first survey (April 1999), contained three questions: (1) did you take any trips to Hueco Tanks in the last twelve months; (2) if yes, how many trips did you take in the last twelve months; and (3) if you took any trips to Hueco Tanks in the last twelve months, what was the average length of stay? This survey was mailed to 387 of the 413 climbers who participated in the first survey (26 of the original 413 surveys were returned as undeliverable). Of this amount, 246 climbers responded, representing a 64% response rate.[1] Table 1 summarizes the trip data collected from each survey instrument.

A construct validity test of scope is employed where pre-policy SP trip data can be compared with PRE-RP and post-policy RP (POST-RP) trip data. The test of scope is conducted by treating Hueco Tanks as a categorically nested good (Carson and Mitchell 1995). Categorical nesting exists when a good G is composed of two or more objects, such as g and its complement g', where neither g nor g' is an empty set and their intersection is empty (Carson and Mitchell 1995). For example, a park area G may be comprised of several areas within the park, where g is a proper subset of those areas. Hueco Tanks is comprised of four separate areas within the park, where access to all areas constitute the good G, and access to some subset of areas would be g. It is always possible to have multiple levels of nests, but to maintain the property of categorical nesting, in each case the lowest category in the nest must be a proper subset of the next higher nest. Table 1 describes the level of access to Hueco Tanks being evaluated, and in each case, the study design maintains the property of categorical nesting.

To assess the construct validity of SP trip responses, a few assumptions are specified. First, recreation demand for Hueco Tanks is a normal good. Second, the level of access at Hueco Tanks is assumed weakly complementary with rock climbing trips. Thus, a climber would not derive utility from an increase in access when climbing trips are zero (i.e., the value of access strictly represents a use value). As an indicator of support for this assumption, <u>all</u> climbers indicated in their completed surveys that the primary purpose of their visit was to climb or boulder. Third, the reservation system implemented in 1998 is not a binding constraint on total

[1] Sample descriptive statistics, such as years experience, climbing ability, demographic, and socioeconomic characteristics, indicate that the sample of climbers participating in both surveys is not statistically different from the sample participating in the 1998 survey only. Further, we could not reject a set of hypotheses that tested whether the average number of PRE-RP trips and SP trips taken (or stated) by both samples were equal.

trips during the season.[2] Finally, the fourth assumption is that respondents get positive utility from using g even after using its complement g'.

Based on these assumptions, the values for different elements of a good G should vary according to the level of inclusion of the good. This means that the value of the good G should be greater than the value of a subset g. Because each level of access to Hueco Tanks being evaluated maintains the property of categorical nesting, it is expected that respondents would value access to more areas in the park higher than access to fewer areas. The hypothesis is that significant changes in site access at Hueco Tanks will cause significant changes in a climber's seasonal consumer surplus (SCS) according to the following relationship:

$$\mathbf{H_1}: SCS_{PRE\text{-}RP} \geq SCS_{SP1} \geq SCS_{SP2} \geq SCS_{POST\text{-}RP}.$$

Testing $\mathbf{H_1}$, or any binary comparison of the ordered relationship, constitutes a test of scope for a categorically nested good. Evidence in support of $\mathbf{H_1}$ would be evidence of construct validity. Further, a criterion validity test can also be conducted by comparing SP to RP trip data; note that the PRE- and POST-RP access conditions bound the two SP cases.

Count Data Travel Cost Models

In testing $\mathbf{H_1}$, several single site (i.e. Hueco Tanks) travel cost demand models are estimated. In specifying the demand function, RP and SP trip data are pooled in a single model. An advantage of pooling RP and SP trip data is that the researcher can test for differences in empirical results derived from different sources of data (Eiswerth et al. 2000). In addition, use of a single site travel cost demand model is likely a defensible approach when the site is relatively unique (Eiswerth et al. 2000). Indeed Hueco Tanks is a unique climbing resource known to climbers throughout the world; climbers often indicated in their completed surveys that no substitute site for Hueco Tanks exists.

A pooled travel cost model can be represented by:

$$v_{ijt} = f\left(tc_{ij}, y_{ij}, sd_i, q_{ijt}, D_t, D_t \times tc_{ij}\right), \tag{1}$$

where v_{ijt} is the number of observed or intended visits that individual i took to site j under access conditions t, tc_{ij} is the travel cost for individual i to site j, y_{ij} is the income available to individual i on their visit to site j, sd_i is a vector of socioeconomic characteristics of individual i, q_{ijt} is a vector of site characteristics experienced by individual i at site j under access conditions t, D_t is a dummy variable indicating the access conditions (i.e., SP1, SP2, POST-RP), and $D_t \times tc_{ij}$ is the interaction of dummy variables and travel costs.

A number of count data econometric techniques have been applied to travel cost models of recreation demand (Eiswerth et al. 2000; Englin and Cameron 1996; Rosenberger and Loomis 1999; Shaw and Jakus 1996). While several econometric techniques can be applied to count

[2] An anonymous reviewer raised a concern that the introduction of the park reservation system in 1998 might be acting as an additional rationing mechanism on visitation behavior. Using total reservation and visitation data provided by TPWD for the period, we determined that the reservation system did not impose a binding constraint on total trips for the season. Based on daily averages, the number of walk-in visits to North Mountain exceeded the number of reserved visits by 27%. On days in which there were 50 visitors at North Mountain (records show that this occurred 40 out of 259 days), users had other options for entering the park, such as waiting for a visitor to leave or entering a different area of the park by guided tour.

data, this study employs pooled Poisson and Negative Binomial (NB) regression models. Pooled Poisson or NB regression models can be estimated if the systematic variation across demand equations is captured by independent variables.[3]

The Poisson regression model assumes that v_{ijt}, given a vector of regressors $\mathbf{x}_i$ defined in equation (1), is independently Poisson distributed with density (Cameron and Trivedi 1998, 20)

$$(2)\qquad f\left(v_{ijt}|\mathbf{x}_i\right)=\frac{e^{-\lambda_i}\lambda_i^{v_{ijt}}}{v_{ijt}!}\qquad v_{ijt}=0,1,2,\ldots,N_i \text{ trips}$$

and mean parameter specified as an exponential link function:

$$(3)\qquad \lambda_i=\exp\left(\mathbf{x}_i'\beta\right),$$

where β are the vector of parameters to be estimated. The exponential link function ensures that the parameter λ_i is nonnegative. Further, the Poisson regression model assumes that the conditional mean, $E[v_{ijt}|\mathbf{x}_i]$, and variance $V[v_{ijt}|\mathbf{x}_i]$ are equal (i.e., equidispersed). The log-likelihood function, maximized over n individuals for the Poisson regression model is

$$(4)\qquad \ln L=\sum_{i=1}^{n}\left[-\lambda_i+v_{ijt}\left\{f\left(tc_{ij},y_{ij},sd_i,q_{ijt},D_t,D_t\times tc_{ij}\right)\right\}-\ln\left(v_{ijt}!\right)\right].$$

For count data models emphasis is often placed on the assumption of the correct specification of the conditional mean and variance (Cameron and Trivedi 1998). As an alternative to Poisson, one can specify a distribution that permits more flexible modeling of the variance by relaxing the assumption that the variance equals the mean, yet maintains the assumption that the mean is $\exp(\mathbf{x}_i'\beta)$. In this framework, a gamma-distributed unobserved individual heterogeneity term is introduced in the Poisson model to take account of dispersion in the data (Cameron and Trivedi 1998, 71). Following Cameron and Trivedi (1998, 63), the NB variance, ω_i, is specified as a general variance function of the mean and dispersion scale parameter α:

$$(5)\qquad \omega_i=\lambda_i+\alpha\lambda_i^{2-\kappa}.$$

The κ parameter allows the relation between the conditional mean and variance to take a variety of forms. For $\kappa = 1$, the variance is specified as a linear function of the mean; this specification is referred to as the NB1 variance function. The NB2 variance function sets $\kappa = 0$, where the variance is quadratic in the mean. In both the NB1 and NB2 the dispersion parameter α is to be estimated. In a generalized NB (GNB) model both α and κ are estimated.

[3] As an alternative, Englin and Cameron (1996) suggest using panel data methods for RP and SP data to handle unobserved individual heterogeneity not captured by explanatory variables. In their study, Englin and Cameron (1996) estimated fixed effects Poisson regression models. Similarly, Rosenberger and Loomis (1999) apply a random effects Poisson regression model to value ranchland to tourists visiting a resort town in the Rocky Mountains.

The NB regression models are estimated by maximum likelihood. For an independent sample of n individuals, the log-likelihood function for NB models is

(6) $$\log L = \sum_{i=1}^{n}\left[\ln\left[\Gamma\left(\zeta_i + v_{ijt}\right)\right] - \ln\left[\Gamma\left(\zeta_i\right)\right] - \ln v_{ijt}! + \zeta_i \ln\left(\frac{\zeta_i}{\zeta_i + \lambda_i}\right) + v_{ijt}\ln\left(\frac{\lambda_i}{\zeta_i + \lambda_i}\right)\right],$$

where

(7) $$\zeta_i = \alpha^{-1}\lambda_i^{\kappa}.$$

When $\kappa = 1$ and $\kappa = 0$ equation (6) simplifies to the log-likelihood function for the NB1 and NB2 regression models, respectively.

In testing $\mathbf{H_1}$, estimates for SCS need to be calculated. Following Bockstael et al. (1984) and given the set of interaction terms ($D_t \times tc_{ij}$), the estimated individual SCS for each policy scenario can be calculated as:

(8a) $$SCS_{PRE-RP} = \frac{-\hat{v}_{it}}{\beta_{TC(PRE-RP)}},$$

(8b) $$SCS_{SP1} = \frac{-\hat{v}_{it}}{\left(\beta_{TC(PRE-RP)} + \beta_{TC(SP1)}\right)},$$

(8c) $$SCS_{SP2} = \frac{-\hat{v}_{it}}{\left(\beta_{TC(PRE-RP)} + \beta_{TC(SP2)}\right)},$$

(8d) $$SCS_{POST-RP} = \frac{-\hat{v}_{it}}{\left(\beta_{TC(PRE-RP)} + \beta_{TC(POST-RP)}\right)},$$

where $\hat{v}_{it}$ is the predicted number of trips taken by individual i under access conditions t. The term $\beta_{TC(\text{PRE-RP})}$ is the estimated coefficient on the base category of travel costs, and $\beta_{TC(\text{SP1})}$, $\beta_{TC(\text{SP2})}$, and $\beta_{TC(\text{POST-RP})}$ are the estimated coefficients on the interaction of data source dummy variables and travel costs.

Dependent and Explanatory Variables

For pooled Poisson and NB regression models the dependent variable, v_{ijt}, is comprised of PRE-RP, SP1, SP2, and POST-RP trip data. SP1 and SP2 intended trip data are constructed by adding (subtracting) the increase (decrease) in intended visitation to PRE-RP trips. After eliminating surveys with inconsistent or missing contingent behavior responses, the number of observations for PRE-RP, SP1, and SP2 trip data is 390. For POST-RP trip data the number of observations is 239. The mean number of trips for each data source is presented in Table 1. For ALL TRIPS (i.e., all trip data response sources combined), the size of the standard deviation-to-mean ratio of 2.575 is an indication of overdispersion (i.e., a variance in excess of mean), possibly resulting from a large number of zero observations in SP2 and POST-RP data sources. Approximately, 42 percent of climbers stated that they would not take any trips if both East and West Mountain were closed (i.e., 165 zero observations out of 390). By comparison, 199 out of 239 climbers (83 percent) did not take any trips to Hueco Tanks after the change in policy. Overall, 35 percent of 1409 observations are zero trips.

Explanatory variables used in Poisson and NB models are shown in Table 2. The independent variables include: travel costs (TC); the number of boulder problems available under various policy site access rules (BPROBLEM); whether a climber spent most of their time at Hueco Tanks climbing at North Mountain (NORTH); whether a climber had knowledge regarding TPWD's intent to propose a climbing management plan (KNOW); dummy variables denoting site access conditions (i.e., DUMSP1, DUMSP2, and DUMPOST); interaction terms between dummy variables and other explanatory variables; socioeconomic variables; and indicators of climber experience and type. Further, because some climbers with knowledge about proposed management plans may have had an incentive to influence outcomes, KNOW is interacted with SP trip responses to control for strategic responses.

Because we are primarily interested in testing the validity of CB data, we use a rather conservative specification of TC. Travel expenditures are calculated as the product of an individual's per mile travel expense and their roundtrip travel miles. In this study, \$0.325 is used for per mile travel expenses.[4] The shortest road distance in miles between two zipcodes is calculated using ZIPFIP (Hellerstein et al. 1993).

It is also argued that the number and difficulty of boulder problems or climbs available at a site (BPROBLEM) will influence a climber's demand for climbing at Hueco Tanks. Because it is believed that climbers select areas at Hueco Tanks that offer boulder problems comparable with their skills, the variable BPROBLEM is constructed to take into account climber skill differences and site characteristics; thus, BPROBLEM is a continuous variable that measures changes in site access.

The dummy variables and the interaction of these variables with travel costs are included in the pooled count data regression models to measure changes in consumer surplus. It is hypothesized that major changes in site access should result in significant differences in parameter estimates on travel costs, and thus different estimates of consumer surplus (CS).

Empirical Results

The results from the GNB model are presented in Table 3. Evidence from t-tests and likelihood ratio (LR) tests indicate that the GNB model (see bottom section of Table 3) is favored over NB1 and NB2 models. Further, perhaps because of the number of zero trip observations associated with greater restrictions in site access, the Poisson model is also rejected. Thus, Table 3 reports results for the pooled GNB model only.

The GNB model performs well with a number of estimates significant at the 0.01 level. The coefficient on YRCLIMB is negative and significant at the 0.01 level, while the coefficient on the quadratic term of YRCLIMB ($YRCLIMB^2$) is positive and significant at the 0.05 level; thus suggesting a U-shaped relationship exists between the number trips and years climbing experience. Because Hueco Tanks is primarily a bouldering area and climbers generally do not need to be skilled in climbing protective gear (a skill generally associated with years of

[4] \$0.325 is the standard mileage rate allowed by the Internal Revenue Service for 1998 business travel expense deductions. This amount takes into account basic car expenses including depreciation, maintenance and repairs, gasoline, oil, insurance, and vehicle registration fees.

experience), these results are not surprising. Further, results show that those climbers who consider themselves boulderers (BOULD) will take more trips to Hueco Tanks. Socioeconomic variables that affect visits are MALE and HH; the coefficient on MALE is negative and significant at the 0.10 level and the coefficient on HH is positive and significant at the 0.05 level.

Overall the estimated coefficients on TC, BPROBLEM, NORTH, KNOW, and TCP are strong determinants of trip-taking behavior to Hueco Tanks; the coefficients on these site specific variables are significant at the 0.01 level. The estimated coefficient on TC is negative as expected. The number of boulder problems (BPROBLEM) available to an individual (depending on site characteristics and climber ability) has an expected positive sign. The estimated coefficients on NORTH, KNOW, and TCP are positive, suggesting that climbers who prefer North Mountain, who had prior knowledge regarding the possibility of site closure, and who owned a TCP were likely to take more trips to Hueco Tanks.

The coefficients of interest are those on the dummy variables and the interaction of the dummy variables with TC and KNOW. The estimated coefficients on DUMSP2 and DUMPOST are positive and significant at the 0.01 level. The coefficients on DUMSP2×TC and DUMPOST×TC are negative and highly significant at the 0.01 level; these results suggest that climber trip behavior changes significantly when an increasing number of areas at Hueco Tanks are closed. Further, it appears that climbers did not give statistically different behavioral responses to CB questions based on their prior knowledge of the possibility of site closure. (The coefficients on DUMSP1×KNOW and DUMSP2×KNOW are not statistically significant.)

To test $\mathbf{H_1}$, a Wald test is conducted to explore differences in parameter estimates. A Wald test provides the appropriate hypothesis test for differences in trip behavioral responses because of the consistency of the covariance matrix (Gourieroux et al. 1984). The null hypothesis is the following set of independent restrictions:

$\mathbf{H_2}$: $\beta_{TC} = (\beta_{DUMSP1\times TC} + \beta_{TC})$
$\mathbf{H_3}$: $\beta_{TC} = (\beta_{DUMSP2\times TC} + \beta_{TC})$
$\mathbf{H_4}$: $\beta_{TC} = (\beta_{DUMPOST\times TC} + \beta_{TC})$
$\mathbf{H_5}$: $\beta_{DUMSP1\times TC} = \beta_{DUMSP2\times TC}$
$\mathbf{H_6}$: $\beta_{DUMSP1\times TC} = \beta_{DUMPOST\times TC}$
$\mathbf{H_7}$: $\beta_{DUMSP2\times TC} = \beta_{DUMPOST\times TC}$

The set of hypotheses tests determine if visitation data exhibit statistically significant differences across substantial changes in site access. Results of these hypotheses tests are listed in Table 4.

The estimated coefficients on travel costs are –2.73 (β_{TC}), –2.81 ($\beta_{TC} + \beta_{DUMSP1\times TC}$), –3.94 ($\beta_{TC} + \beta_{DUMSP2\times TC}$), and –8.78 ($\beta_{TC} + \beta_{DUMPOST\times TC}$). According to hypotheses tests $\mathbf{H_3}$ and $\mathbf{H_5}$, closure of both East and West Mountain—a 43 percent reduction in boulder problems—leads to statistically different estimates at the 0.01 level. Hypotheses tests $\mathbf{H_4}$, $\mathbf{H_6}$ and $\mathbf{H_7}$ test whether the most restrictive policy results in statistically different estimates; $\mathbf{H_4}$ shows that $\beta_{DUMPOST\times TC}$ is statistically different from zero; $\mathbf{H_6}$ show that $\beta_{DUMPOST\times TC}$ is statistically different from $\beta_{DUMSP1\times TC}$; and $\mathbf{H_7}$ shows that $\beta_{DUMPOST\times TC}$ is statistically different than $\beta_{DUMSP2\times TC}$. Hypothesis $\mathbf{H_2}$ that the coefficients β_{TC} and $\beta_{TC} + \beta_{DUMSP1\times TC}$ are equal is not rejected. Failure to reject $\mathbf{H_2}$ is not surprising, however, because SP1 is associated with the closure of West Mountain that consists of only 9 percent of available boulder problems. Overall, these results indicate that CS will be statistically different across <u>major</u> differences in site closures at Hueco Tanks.

Estimates of per trip CS and SCS are presented in Table 5. For example, per trip CS is \$366 for access to four mountain areas versus \$114 for access to one mountain area. Given the uniqueness of Hueco Tanks, these per trip CS measures seem reasonable. Further, as hypothesized in $\mathbf{H}_1$, SCS measures get increasingly smaller as more sites are closed at Hueco Tanks. The average seasonal loss to climbers due to restricted access to two areas (East and West Mountain) is \$687 per climber, and \$1276 per climber when three areas are closed (the actual policy change). Thus, the values for access to mountain areas in Hueco Tanks are sensitive to scope.

A concern with the results, however, may be due to the one-year period between surveys, in which climate conditions may have differed. It is typically expected that higher temperatures and lower precipitation are positively correlated with trips. Climate data was not included in the regressions because this data does not vary across individuals. Average temperatures in El Paso during the POST-RP period were 2.27° Fahrenheit higher and precipitation was 0.15 inches lower than the PRE-RP period.[5] Thus, although climate conditions were favorable for climbing trips during the POST-RP period, it appears that any resulting increases in trips did not outweigh decreases in trips caused by the access restrictions.

In the case of rock climbing at Hueco Tanks, results from this study support the validity of CB trip data, and do not suggest that climbers overstate changes in their trip-taking behavior in CB responses. Climbers are able to project the direction of their behavioral response to area closures in a way that is consistent with economic theory; in this case, climbers demand for climbing at Hueco Tanks decreases as more areas are closed within the park.

Conclusions

In 1998, TPWD restricted open-recreational access at Hueco Tanks State Park, a world-class climbing site. The implementation of this policy change provided a unique opportunity to test the validity of contingent behavior data.

To collect trip data from climbers before and after the policy change, two separate surveys were implemented. The first survey was conducted prior to the restriction in open-recreational access. In this survey, climbers provided information about their pre-policy rock climbing trips to Hueco Tanks and intended trips under alternative, hypothetical policy rules restricting access. The second survey was conducted after the restriction on access was imposed; climbers were surveyed about their post-policy rock climbing trips.

A construct validity test of scope was conducted by comparing post-policy RP trip data with pre-policy RP and CB data. Results from a generalized Negative Binomial regression model indicate that climbers do not appear to overstate changes in trip behavior when presented with hypothetical questions about site access restrictions. In addition, for major decreases in site access, climbers' values for Hueco Tanks also significantly decrease. Together, these results support the conclusion that CB trip response data are sensitive to changes in scope. Thus, methods of augmenting RP data sets with SP data show promise as a tool for estimating demand, and as an input for public land management decisions.

References

American Psychological Association. 1974. *Standards for Educational and Psychological Tests*. Washington, D.C.

[5] Data provided by El Paso Climate Directory on the World Wide Web at http://nwselp.epcc.edu/elp/wxclim.html.

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

Berrens, R.P., and R.M. Adams. 1998. "Applying Contingent Valuation in the Design of Fee Hunting Programs: Pheasant Hunting in Oregon Revisited." *Human Dimensions of Wildlife* 3(3):11-25.

Bockstael, N., W.M. Hanemann, and I.E. Strand. 1984. *Measuring the Benefits of Water Quality Improvements Using Recreational Demand Models*. Vol. 2 of *Benefits Analysis using Indirect or Imputed Market Methods*. EPA contract number: CR-811043-01-0.

Cameron, A.C., and P.K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Carson, R.T. 1997. "Contingent Valuation Surveys and Tests of Scope Insensitivity." In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, eds. R.J. Kopp, W. Pommerhene, and N. Schwartz. Boston: Kluwer.

Carson, R.T., N.E. Flores, and R.C. Mitchell. 1999. "The Theory and Measurement of Passive-Use Value." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, eds. I.J. Bateman and K.G. Willis. New York: The Oxford University Press.

Carson, R.T., and R.C. Mitchell. 1995. "Sequencing and Nesting in Contingent Valuation Surveys." *Journal of Environmental Economics and Management* 28:155-173.

Cicchetti, C.J., and N. Peck. 1989. "Assessing Natural Resources Damages: The Case Against Contingent Value Survey Methods." *Natural Resources and Environment* 4(1).

Eiswerth, M.E., J. Englin, E. Fadali, and W.D. Shaw. 2000. "The Value of Water Levels in Water-Based Recreation: A Pooled Revealed Preference/Contingent Behavior Approach." *Water Resources Research* 36(4):1079-86.

Englin, J. and T.A. Cameron. 1996. "Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7:133-47.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* 52(3):701-20.

Greene, W. 1997. *Econometric Analysis*. New Jersey: Prentice Hall.

Hellerstein, D., D. Woo, D. McCollum, and D. Donnelly. 1993. *ZIPFIP: A Zip and FIPS Database*. Washington D.C.:U.S. Department of Agriculture, ERS-RTD.

Hueco Tanks State Historic Park. 1997. *Hueco Tanks SHP Resource Management Plan* draft 3.2.

Loomis, J. 1999. "Recreation and Passive Use Values From Removing the Dams On the Lower Snake River to Increase Salmon. Report from Agricultural Enterprises, Inc. to U.S. Army Corp of Engineers. Walla Walla, WA.

Loomis, J. 1993. "An Investigation into the Reliability of Intended Visitation Behavior." *Environmental and Resource Economics* 3:183-191.

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

Nestor, D. 1998. "Policy Evaluation with Combined Actual and Contingent Response Data." *American Journal of Agricultural Economics* 80:264-276.

Rosenberger, R., and J. Loomis. 1999. "The Value of Ranch Open Space to Tourists: Combining Observed and Contingent Behavior Data." *Growth and Change* 30(summer):366-83.

Shaw, W.D. and P. Jakus. 1996. "Travel Cost Models of the Demand for Rock Climbing." *Agricultural and Resource Economics Review* 25(2):133-142.

Texas Parks and Wildlife Department. 1998. *Hueco Tanks State Historical Park Public Use Plan*. Austin, TX.

Texas Parks and Wildlife Department. 1997. *Hueco Tanks State Historical Park Draft Public Use Plan*. Austin, TX.

U.S. National Park Service (NPS). 1993. *Federal Register* 58(112):32878-80.

Vatn, A. and D. Bromley. 1995. "Choices Without Prices Without Apologies." In *Handbook of Environmental Economics*, ed. D. Bromley. Cambridge: Blackwell Publishers, Inc.

Vossler, C.A., and J. Kerkvliet. 1999. "A Nonexperimental Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior." Manuscript, Department of Agricultural, Resource, and Managerial Economics, Cornell University.

Table 1: Trip Data Collected From Survey Instruments

Date of Survey Instruments	Data Collected	Acronym	Site Access Conditions	Number of Available Boulder Problems	Mean Number of Trips (Standard Deviation) [number of observations]
1998 Survey	Pre-policy revealed preference climbing trip data	PRE-RP	North Mountain East Spur Maze East Mountain West Mountain	1237	5.487 (11.946) [390]
	Contingent behavior climbing trip data to hypothetical changes in site access	SP1	North Mountain East Spur Maze East Mountain	1127	5.226 (11.841) [390]
		SP2	North Mountain East Spur Maze	706	3.867 (10.838) [390]
1999 Survey	Post-policy revealed preference climbing trip data	POST-RP	North Mountain	509	1.335 (6.623) [239]
	All trip data sources combined	ALL TRIPS	—	—	4.262 (10.975) [1409]

Table 2: Description of Independent Variables

Variable	Description	Mean (Standard Deviation)
PREDUM	Dummy variable – 1 indicates if data is PRE-RP, 0 otherwise. This is the base category dropped during estimation.	
DUMSP1	Dummy variable – 1 indicates if data is SP1, 0 otherwise.	0.277
DUMSP2	Dummy variable – 1 indicates if data is SP2, 0 otherwise.	0.277
DUMPOST	Dummy variable – 1 indicates if data is POST-RP, 0 otherwise.	0.169
YRCLIMB	Number of years climbing experience. Variable scaled by 100.	0.076 (0.066)
$YRCLIMB^2$	YRCLIMB squared.	0.010 (0.018)
BOULD	Dummy variable – 1 indicates whether the person primarily is a boulderer, 0 otherwise.	0.087 (0.281)
TC	Roundtrip travel miles at $0.325 per mile. Variable scaled by 1000	0.482 (0.385)
BPROBLEM	Number of boulder problems available at Hueco Tanks based on different site access conditions and a climber's ability range using the "V" rating system. Variable scaled by 1000.	0.731 (0.306)
NORTH	Dummy variable – 1 indicates that a climber spent the majority of her time climbing and bouldering at North Mountain, 0 otherwise.	0.219 (0.413)
KNOW	Dummy variable – 1 indicates the climber had information prior to taking a trip regarding the intent of the TPWD to propose a climbing management plan for Hueco Tanks, 0 otherwise.	0.661 (0.473)
TCP	Dummy variable – 1 indicates the climber owned a Texas Conservancy Passport, which allowed them to enter Hueco at a user fee discount, 0 otherwise.	0.222 (0.416)
UNEARNY	The annual amount of a climber's unearned income scaled by 10000.	0.338 (0.992)
MALE	Dummy variable – 1 indicates the climber is male, 0 otherwise.	0.789 (0.408)
TOTHOURS	Total hours an individual worked during the year scaled by 10000.	0.158 (0.078)
HH	Number of members in climber's household.	2.137 (1.184)
PEOPLE	Average number of people traveling with climber to Hueco Tanks (including the climber)	3.140 (2.204)

Table 3: Parameter Estimates for GNB Regression Model

GNB Regression Model			
Intercept	0.246	NORTH	0.384***
	(1.14)[a]		(4.46)
DUMSP1	0.201	KNOW	0.406***
	(1.33)		(3.58)
DUMSP2	0.712***[b]	DUMSP1×KNOW	-0.077
	(3.99)		(-0.48)
DUMPOST	0.618***	DUMSP2×KNOW	-0.044
	(2.91)		(-0.25)
YRCLIMB	-10.427***	TCP	0.719***
	(-4.92)		(8.67)
YRCLIMB^2	19.971**	UNEARNY	0.022
	(2.43)		(0.60)
BOULD	0.542***	MALE	-0.181*
	(3.95)		(-1.92)
TC	-2.733***	TOTHOURS	0.450
	(-16.53)		(1.14)
DUMSP1×TC	-0.077	HH	0.054**
	(-0.35)		(1.97)
DUMSP2×TC	-1.204***	PEOPLE	0.024
	(-4.29)		(1.36)
DUMPOST×TC	-6.050***	α	0.400***
	(-7.21)		(5.69)
BPROBLEM	1.856***	κ	-0.477***
	(9.17)		(-4.70)
Number of Observations	1409		
LnL	-2592.942		
Likelihood Ratio Index[c]	0.344		
Selection Test from the GNB Model[d]	t-test for Poisson [H_0: $\alpha = 0$] t = 5.69***		
	t-test for NB1 [H_0: $\kappa = 1$] t = –5.15***		
	t-test for NB2 [H_0: $\kappa = 0$] t = –4.70***		
	LR test for NB1 [H_0: $\kappa = 1$] $\chi^2 = 464.96$***		
	LR test for NB2 [H_0: $\kappa = 0$] $\chi^2 = 31.18$***		

[a] Numbers in parentheses are the ratio of the estimated coefficient to the aysmptotic standard error.

[b] ***, **, and * denote the estimate is significantly different than zero at the 0.01, 0.05, and 0.10 levels, respectively.

[c] The likelihood ratio index for Poisson is defined as $1\text{-}(LnL_{fit}/LnL_{restricted})$, where LnL_{fit} and $LnL_{restricted}$ are the LnL values for the fitted and intercept-only models (Cameron and Trivedi 1998, 155). The likelihood ratio index for the negative binomials is defined as $1\text{-}(LnL_{NB}/LnL_{fit})$, where the subscript NB refers to NB1, NB2, and GNB.

[d] The hypotheses in brackets represent the implied restrictions in the GNB model. The likelihood ratio (LR) test statistic is defined as $-2[LnL_{restricted} - LnL_{GNB}]$, where the restricted model is either NB1 or NB2. The LR test is distributed as χ^2 with ($K_{GNB} - K_{restricted}$) degrees of freedom, where K refers to the numbers of estimated coefficients in each model.

Table 4: Hypotheses Tests for Validity

Hypothesis Test	Description[a]	χ^2
H_2	$\beta_{TC} = \beta_{DUMSP1 \times TC} + \beta_{TC}$	0.047
H_3	$\beta_{TC} = \beta_{DUMSP2 \times TC} + \beta_{TC}$	9.32***[b]
H_4	$\beta_{TC} = \beta_{DUMPOST \times TC} + \beta_{TC}$	46.82***
H_5	$\beta_{DUMSP1 \times TC} = \beta_{DUMSP2 \times TC}$	15.79***
H_6	$\beta_{DUMSP1 \times TC} = \beta_{DUMPOST \times TC}$	50.50***
H_7	$\beta_{DUMSP2 \times TC} = \beta_{DUMPOST \times TC}$	32.21***

[a] The estimated coefficients on travel costs are -2.733, -2.813, -3.937, -8.783 for β_{TC}, $\beta_{TC} + \beta_{DUMSP1 \times TC}$, $\beta_{TC} + \beta_{DUMSP2 \times TC}$, and $\beta_{TC} + \beta_{DUMPOST \times TC}$, respectively.
[b] *** denotes that the β coefficients are significantly different than each other at the 0.01 level.

Table 5: Consumer Surplus Measures from GNB Model

Policy Change	Per Trip Consumer Surplus	Seasonal Consumer Surplus
Open access to four mountain areas at Hueco Tanks	\$366***[a] (22.13)[b]	\$1640
Open access to three mountain areas at Hueco Tanks (West Mountain closed)	\$355*** (45.02)	\$1553
Open access to two mountain areas at Hueco Tanks (West and East Mountain closed)	\$254*** (25.44)	\$953
Open access to one mountain area (North Mountain) at Hueco Tanks; however, additional access restrictions apply.	\$114*** (11.46)	\$364

[a] *** denotes that the estimate is significantly different than zero at the 0.01 level.
[b] Standard errors in parentheses. These standard errors are calculated using the Delta Method Approximation (Greene 1997, 278).

Is Actual Willingness to Pay for a Public Good Sensitive to the Elicitation Format? Implications for Contingent Values

Patricia A. Champ
USDA Forest Service, Rocky Mountain Research Station
2150 Centre Ave, Suite 350
Fort Collins, CO 80526
970-295-5967 (phone)
970-295-5959 (fax)
champ@lamar.colostate.edu

Richard C. Bishop
University of Wisconsin
Department of Agricultural and Applied Economics
427 Lorch Street
Madison, WI 53706
608-262-8966 (phone)
608-262-4376 (fax)
bishop@aae.wisc.edu

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Acknowledgements: This study was funded by the USDA Forest Service, Rocky Mountain Research Station and the Energy Center of Wisconsin. We are grateful for the cooperation of Greg Bollom and Jeff Ford from Madison Gas and Electric. Mike Welsh generously provided help with the statistical estimation. John McPeak managed the data collection. All remaining errors are the sole responsibility of the authors.

Introduction

Stated preference data are essential for some of the non-market valuation techniques, namely contingent valuation and attribute-based stated preference methods. Likewise stated preferences are also used for assessing the potential market for new voluntary environmental programs. While use of such data is prevalent and essential, some economists remain skeptical about the validity of stated preference data for myriad reasons. One of the more widely researched criticisms particular to contingent valuation data has been the consistent finding that the elicitation format of the willingness to pay (or willingness to accept) question influences the value estimates. The conventional wisdom has been that willingness to pay questions with differing elicitation formats (i.e. open ended, dichotomous choice, payment card, double bounded dichotomous choice, etc) all measure the same latent construct (Hicksian surplus measure) and should therefore provide equivalent value estimates. This conventional wisdom implicitly assumes the measurement biases associated with the various elicitation formats are commensurate. Therefore found elicitation effects have been portrayed as an artifact of the hypothetical nature of the contingent valuation method (McFadden 1994). There has been some questioning of this premise. For example, Bohara, McKee and Berrens (1998) suggest that differences between elicitation formats" may be justified if respondents are actually thinking about their answers and responding to the information provided differently in the two formats." In other words, the elicitation effects are not due to the hypothetical nature of the contingent valuation method but rather are a survey research issue.

In this study we take a step back and investigate whether there are elicitation effects associated with actual payments for a public good. The motivation for investigating elicitation effects with actual payments is that actual payment decisions are what contingent values are intended to mimic and respondents to a contingent valuation survey should behave as they would in an actual payment situation. Therefore if there are elicitation effects when individuals are making real payment decisions, there should be elicitation effects when individuals are making hypothetical payment decisions. If elicitation effects are expected, a more productive research focus may be on trying to understand the causes of elicitation effects and developing a framework for assessing the relative validity of the various elicitation formats.

The study described in this manuscript makes several contributions to the current state of knowledge. First it involves two actual payment treatments, in one the willingness to pay question is posed in the dichotomous choice format and in the other treatment, the willingness to pay question uses a payment card response format. The public good is a voluntary wind power program whose benefits accrue to a population far beyond those that pay for the program. The analysis of the data focuses on both the willingness to pay estimates and multivariate models estimated with consistent maximum likelihood functions. In some of the previous elicitation studies, the statistical methods used to estimate willingness to pay for differing types of data also differed. In those cases, it is possible that found differences were due to *how* the measures were estimated.

In the next section we summarize some of the previous elicitation studies. We also describe the hypotheses we set out to test. Then we describe the data collection effort and the wind power program. In the last sections we summarize the results of our empirical study and provide some discussion.

Background

In the survey research literature there is an extensive discussion of the effects of posing a question with an "open" versus "closed" ended response format (Sheatsley 1983). An "open" response format does not provide any information to the respondent about possible answers to the question while a "closed" format provides answer categories from which a response is chosen. The focus on elicitation effects associated with contingent values has been on open versus closed response formats and on differences between various closed response formats. Two of the most prevalent closed response formats used in contingent valuation are the dichotomous choice and payment card formats.[1] Each of these formats has relative strengths and weaknesses in terms of cognitive burden on the respondent, statistical efficiency, incentive compatibility, and anchoring effects. In this study we compare the payment card format to the dichotomous choice format. The benefits of the dichotomous choice format include an endorsement by the NOAA panel (Arrow et al. 1993) for use in damage assessments, a relatively lower cognitive burden on the respondents as the format replicates real life decision making (purchases in markets and voting), incentive compatibility under some circumstances (Carson, Groves, and Machina 1999). The drawbacks of the dichotomous choice format include the difficulty in choosing the appropriate offer amounts (Alberini 1995) and the need for large sample sizes due to the statistical inefficiency of such data. By comparison the payment card format allows for smaller sample sizes relative to the dichotomous choice format. However, choosing an amount from the payment card may be a more cognitively burdensome task for respondents. There have also been concerns about the effects of the range and increments used on the payment card.[2] The bottom line is that one elicitation format is not unequivocally better than the other. In the case of voluntary environmental programs, the payment card format is more consistent with how actual programs are implemented in that the individual chooses the level at which to participate.

Table 1 summarizes recent elicitation studies. Some of the comparisons implemented different statistical approaches to estimate the willingness to pay values for the various treatments. For example, with open ended data mean willingness to pay is sometimes calculated as the simple mean of the responses while with dichotomous-choice data, willingness to pay estimate are usually based on a parametric models (i.e. logit or probit). It is possible that some of differences

---

[1]The dichotomous choice format poses the willingness to pay question in a 'take-it-or-leave-it' format. The individual responds either *yes* or *no* to paying a specified amount. The payment card format asks which dollar amount from a menu of amounts, is the maximum the individual would be willing to pay.

[2]Rowe et al. (Rowe, Schulze, and Breffle 1996) argue that the range of the dollar values is not an issue if a sufficiently large range is used on the payment card.

in reported results are statistical rather than data generated as argued by Huang and Smith (1998). There are a few notable patterns in the study results. First, in general, the contingent valuation willingness to pay estimates based on closed ended response formats (namely dichotomous choice and payment card) are larger than those based on open ended data. Furthermore, dichotomous choice estimates tend to be larger than payment card estimates. This result is consistent whether the study involved a private or a public good. There are a few cases (Reaves, Kramer, and Holmes 1999; Bohara et al. 1998; Loomis et al. 1997; Boyle et al. 1996; Kealy and Turner 1993) where the willingness to pay estimates were found to be equivalent across elicitation formats. It is also interesting to note that the studies involving actual payments provide very mixed results. Loomis (1997) and Frykblom and Shogren (2000) did not find any difference between willingness to pay estimates for private goods based on open ended and dichotomous choice response formats. One study (Cadsby and Maynes 1999) found willingness to pay estimates for a public good to be larger based on the open-ended data relative to the dichotomous-choice data. This study was a controlled laboratory experiment involving a threshold public good in which the a priori expectation was described as "To the extent that the availability of a symmetric pure strategy equilibrium acts as a behavioral focal point, we should find more contributions and provision in the continuous contribution case which is consistent with such a threshold equilibrium, than in the binary contribution case which is not" (Cadsby and Mynes 1999, p. 57). Actual non-market valuation settings rarely offer the control of the laboratory setting so it is difficult to know how laboratory results would generalize. Two studies (Lunander 1998, Brown et al. 1996) found willingness to pay estimates based on dichotomous choice data to be larger relative to open-ended data. This survey of the literature suggests that the results of the private good experiments are not at all consistent. Of the few actual payment elicitation studies, only Brown et al. and Cadsby and Maynes involved public goods. We might expect elicitation effects to be more prevalent with public goods as respondents are often less familiar and have less experience paying for public goods. In such situations, the elicitation format may send signals to the individual about the value of the good. In particular, the offer amount in the dichotomous choice format might be viewed by some respondents as a signal about the value of the good.

Hypotheses

Before testing the main hypotheses of interest, we will compare the response rates and item non-response for the two treatments. Taking these factors into consideration may allow us to comment on which elicitation format is preferable (Reaves, Kramer, and Holmes 1999). Economic theory provides little guidance on the relationship one would expect between willingness to pay estimates based on differing elicitation formats. If two elicitation formats measure the latent construct without bias or more realistically with an equivalent level of measurement bias (in the same direction), one would expect the two formats to provide equivalent measures of willingness to pay. However, the contingent valuation elicitation effects studies are often motivated by the notion that differing elicitation formats should provide similar estimates of willingness to pay and failure to find equivalent willingness to pay measures is due to the hypothetical nature of the contingent valuation method. In this study we investigate two research questions: 1. Are willingness to pay estimates based on the payment card data and dichotomous choice data equivalent? (H_o: Mean WTP_{pc} = Mean WTP_{dc}) and 2. Are the underlying distributions the same for the responses to the dichotomous choice and payment card willingness to pay question equivalent? (H_0: Dist WTP_{pc} = Dist WTP_{dc}).

The Study Design

To explore the effect of elicitation format on actual payments for a public good, we implemented a split sample design in which the survey instrument and all survey materials were identical for the two treatments except the response format to the willingness to pay question. The good described in the surveys was a voluntary pilot program for residential customers to purchase wind-generated electricity from the Madison Gas and Electric Company (MG&E), a local private provider of gas and electricity in Madison, Wisconsin. The sample frame was MG&E residential customers who were MG&E customers for at least one year.[3] While the program was sponsored by MG&E, the study was conducted by the University of Wisconsin Department of Agricultural and Applied Economics. The data was collected via mail surveys with customers randomly assigned to either the dichotomous choice or payment card treatment.

Survey Development

Eight focus groups with MG&E customers were used to develop and refine all of the survey materials and the description of the wind-power program in particular. Developing the description of the wind-power program was the most challenging aspect of the preliminary design effort. From the first focus groups, it was clear that respondents were a bit confused about what it meant to purchase wind power. Some thought that if they purchased wind power, then the wind power would somehow be delivered directly to their household rather than feeding into the general electricity supply. Based on feedback from the focus groups, we developed a scenario that clarified exactly how the program would work.

Focus groups participants did not actually have an opportunity to participate in the wind-power program because electricity rates must be approved by a regulatory agency, the Wisconsin Public Service Commission, and at the time of the focus groups we did not yet have permission to actually sell wind power. Hence, all focus group participants were asked a contingent donation question. We explored the reasons why focus group participants responded positively to the willingness to pay question. They told us the environmental benefits of wind power and the desire to encourage future renewable energy development were important factors. We also explored reasons for *no* responses to the contingent donation question. Those who said *no* told us that the program was too costly, that all MG&E customers should share in the burden, that it was better to conserve energy than develop new sources, that they objected to the wind generators because of

[3]Madison, Wisconsin is a University town and many of the customers are students who live together in a dwelling for the academic year. Participation in the program required a 12 month commitment so we wanted to avoid students who would be moving sometime during the 12 months. To minimize contacting the student population we also removed multiple-unit dwellings from the sample frame.

potential birds kills, and that they did not believe the wind technology was cost effective. These comments were used to develop a scenario that was balanced in terms of explaining the costs and benefits of wind-generated electricity. Focus group participants also suggested that we make clear in the cover letter why the University of Wisconsin was recruiting customers for an MG&E program. The focus groups also allowed us to test and revise the attitude and demographic questions in the survey. We also conducted a pilot study to allow us to predict the final response rate, provide a test of the survey implementation procedures and assess whether the offer amounts for the payment card and dichotomous choice willingness to purchase questions were appropriate.

The Final Scenario and Valuation Question

The survey explained that MG&E had an opportunity to purchase a limited amount of power from wind turbines being built in northeastern Wisconsin. If enough customers were willing to pay the extra cost of wind power, MG&E would purchase the wind power from the wind turbines and would ultimately construct more wind turbines. Customers were told that most of the electricity MG&E currently sells is generated from coal and the environmental drawbacks of coal were explained. The benefits of wind-generated electricity were described as well as the disadvantages (the sound of the turbines, bird kills, appearance of the wind turbines themselves). It was made clear in the wind program description that the environmental impacts of this specific program would not be substantial; however, if the program were successful MG&E would offer more wind power in the future. The unusual aspect of this good is that payment is linked to a market good (electricity) but the benefits accrue to all customers. In other words, a market mechanism was used to facilitate donations toward the purchase of a public good.

Following the informational part of scenario, survey respondents were asked if they would purchase wind-generated electricity. This is the one difference between the two treatments. In one treatment a single-bound dichotomous choice format was used for this question. Respondents were asked about their willingness to pay a randomly assigned additional amount on their monthly electric bill for one year. The amount they were asked to purchase was independent of their current electricity usage. In the payment card treatment, the respondents were asked to choose the amount of wind-generated electricity they would like to purchase from a menu of amounts. For both treatments the amounts were expressed in kilowatt-hours per month, cost per month, and total cost for the year. The cost per kilowatt-hour was four cents and offer amounts started at 50 kilowatt-hours per month. The next amount was 100 kilowatt-hours per month and the amount increased in 100 kilowatt-hour increments up to 600 kilowatt-hours per month (that corresponds to \$24, \$48, \$96, \$144, \$192, \$240, or \$288 per year). The additional cost of wind-generated electricity is approximately 4.4 cents per kilowatt-hour (residential customers currently pay about 7.4 cents per kilowatt-hour so the total cost of wind-generated electricity is approximately 11.8 cents per kilowatt-hour). The offer amounts used in the dichotomous choice treatment were the same offer amounts listed on the payment card. If survey respondents agreed to purchase the wind-generated electricity, their monthly electricity bill would include the extra cost of wind-generated electricity for twelve months. The actual text of the willingness to participate questions is shown in Figure 1. The format of the willingness to purchase question is different from the traditional willingness to pay format where the quantity of the good is the same across offer amounts. In this case, paying more was linked to purchasing more wind power. However the total amount of wind power that could be purchased was fixed at 430,000 which is a small amount of wind power even for this pilot study population. Individuals who choose to buy more wind power are getting the same benefits as those who choose to purchase less wind power or those who choose to purchase no wind power at all so in that aspect this good is similar to a standard threshold public good. The surveys for both treatments also elicited measures of attitudes toward wind-generated electricity, MG&E, renewable energy, the environment in general and demographic characteristics. These measures allow us to conduct multivariate analyses.

Results

In this section, we first compare the response rates and item non-response for the two treatments then we focus on the responses to the willingness to participate question. The Reaves, Kramer, and Holmes (1999) study is most similar to this study in terms of looking at a public good, comparing payment card and dichotomous choice treatments, and implementing the study via mail surveys. However in the Reaves, Kramer and Holmes study, all treatments were contingent valuation. Throughout this results section, we compare our findings to those of Reaves, Kramer and Holmes.

Response Rate

The initial sample sizes were 1260 for the dichotomous-choice treatment and 500 for the payment-card treatment. Response rates were 56% and 50% respectively giving final samples sizes of 700 for the dichotomous-choice treatment and 251 for the payment-card treatment. The response rate for the dichotomous-choice treatment was significantly higher relative to the payment card treatment ($\chi^2 = 4.134$, $p = .042$). This result is consistent with the premise that the ease with which respondents can answer a dichotomous choice willingness to participate question manifests in more individuals returning the survey. However, we do not really know the cause of the difference in response rate. Reaves, Kramer and Holmes (1999) also compared response rates for payment card and dichotomous choice contingent valuation treatments and found that the response rate for the payment card (50.60%) treatment to be higher relative to the dichotomous choice treatment (46.53%) but the difference was not statistically significant.

Because the response rates were significantly different for the two treatments in this study, analyses were conducted to assess whether the respondents in the two treatments had statistically similar responses to the measures elicited in the surveys other than the WTP question. Based on these analyses we conclude that the two treatments are representative of similar respondent populations.

Item Non-response

One might also expect that more survey respondents would skip (not answer) the willingness to participate question in the payment card treatment if indeed responding to that format is more difficult than responding to a dichotomous choice question. However, that was not observed in this study. Of the 700 surveys returned in the dichotomous choice treatment, 648 had valid responses to the willingness to participate question (i.e. 7% of the returned surveys did not have any response circled for the willingness to participate question). In the payment card treatment, twelve of the 249 returned surveys (5%) did not have a response to the willingness to participate question. The difference in the item non-response for the two treatments is not statistically significant ($\chi^2 = 1.98$; $p = .159$). This result is consistent with the direction of the difference found by Reaves, Kramer and Holmes (1999) in their comparison of the payment card and dichotomous choice contingent valuation treatments. However in their study, the dichotomous choice question was double bounded and the difference in item non-response to the willingness to pay question was statistically significant when comparing the payment card treatment to the double bounded dichotomous choice response. But the difference was *not* statistically significant when comparing the payment card treatment item non-response to non-response to the first dichotomous choice willingness to pay question. In this study, the higher (but not significantly) item non-response in the dichotomous-choice treatment might have been due to survey respondents simply skipping the willingness to participate question and completing the rest of the survey. Whereas in the payment card treatment, it is possible that the some respondents started the survey but threw the entire survey away (thus the lower overall response rate) when they got to the willingness to participate question rather than skipping that question and completing the rest of the survey. Again we can not comment on *why* individuals skip questions or do not return the surveys but it is clear that we got a higher percent of the surveys returned in the dichotomous choice treatment and did not get a significantly higher percent skipping the willingness to participate question.

The aggregate non-response can be calculated by starting with all the individuals who were sent a survey (1260 for the dichotomous treatment and 500 for the payment card treatment) and considering how many of those individuals provided a valid response to the willingness to purchase question (648 for the dichotomous choice and 237 for the payment card). We find no statistical difference between the dichotomous choice and payment card treatments in terms of aggregate non-response ($\chi^2 = 2.324$, $p = .127$). This result suggests that we cannot make a claim based on response rate and item non-response that one treatment is unequivocally better or worse than the other.

Response to the Willingness to Participate Question

When comparing responses to the willingness to participate question between treatments, it is important to keep in mind that the questions are fundamentally different. In the dichotomous choice treatment, the question is whether the individual will participate at the specified payment level. Whereas in the payment card treatment, the question asks if the individual would participate at any one of the payment levels offered on the payment card. This difference leads us to expect more individuals to respond positively to the willingness to participate question in the payment card treatment as the respondent can choose the payment level. We found that to be the case with significantly higher percent of the respondents agreeing to pay for the wind-generated electricity in the payment card treatment (36%) relative to the dichotomous-choice treatment (23%) ($\chi^2 = 14.29$; $p < .001$). However respondents in the dichotomous-choice treatment agreed to purchasing larger quantities of wind power relative to respondents in the payment card treatment as is evident by looking at the empirical survival functions (Table 2 and Figure 2).

The willingness to pay models based on the payment card and dichotomous choice data were estimated using a maximum likelihood interval approach (Welsh and Poe 1998; Cameron and Huppert 1991). Using consistent statistical modeling for the two types of data allows for statistical comparisons between the treatments. The univariate logit models (Table 4) suggest that the underlying distributions differ significantly between the payment card and dichotomous choice treatments. A likelihood ratio test confirms that the distributions are significantly different (LR=35; $\chi^2_{\alpha=.05;2} = 5.99$). Not surprisingly, the estimates of mean willingness to purchase wind-generated electricity based on the logistic models are also very different for two treatments with the mean for the dichotomous choice treatment ($81) more than double that of the payment card treatment ($36).

The multivariate models (Table 5) also suggest that the underlying distributions of willingness to pay are significantly different for the two treatments (LR=27; $\chi^2_{\alpha=.05;9}$=16.93). However we note that with the exception of three coefficient estimates, the other estimated coefficients have similar signs and levels of significance for the dichotomous-choice and payment-card models. One of the differences in the significance level, but *not* in the sign, of the estimated coefficients for the two models occurs for the variable "I like the idea of wind power." For this variable we find that the sign to be positive suggesting individuals who agree or strongly agree with this statement are more likely to actually purchase wind power. However the coefficient is not significant in the payment-card model. A second notable difference between the two multivariate models is found with the variable "I would rather encourage conservation of electricity than pay to develop renewable energy sources." For this variable, individuals who agreed or strongly agreed were less likely to actually purchase wind power. However, the estimated coefficient is significant in the payment-card model but not the dichotomous-choice model. A third disparity between the payment-card and dichotomous-choice multivariate models occurs in the statistical significance of the coefficient on the gender variable in the dichotomous-choice model that is not found in the payment-card model. However the sign on the gender variable in both models suggests that females are more likely than males to purchase wind power. In both models we found the income variable to not be significant. In the models shown in Table 5, the income variable is continuous. This result suggests that support for wind power is independent of income. Written comments in some of the surveys suggested that there are less expensive alternative energy sources, namely nuclear energy. The inclusion of the variable "I can't afford to pay the extra cost of wind power" in the multivariate model may also be a proxy for income. Overall, comparisons of the multivariate models suggest that the differing response formats are consistent with differing underlying preferences.

Conclusions

This study offered a unique opportunity to observe elicitation effects in an actual payment situation for a public good. We argue that elicitation effects associated with contingent values are not likely a result of the hypothetical nature of the willingness to pay questions. Rather elicitation effects are a prevalent survey research issue. As such, the bigger question is which format is most "appropriate" for a particular application?

Table 1: Recent Elicitation Studies since 1990

Authors	The Good	Public or Private Good	Response Formats	Results
Actual Payment Studies				
Frykblom and Shogren (2000)	A Swedish national atlas	Private	DC, OE	$WTP_{DC} = WTP_{OE}$[4]
Cadsby and Maynes (1999)	Tokens which are converted into Canadian Dollars	Public	DC, OE	$WTP_{DC} < WTP_{OE}$
Lunander (1998)	Preview of a movie	Private	DC, OE	$WTP_{DC} > WTP_{OE}$
Loomis et al. (1997)	Art Print	Private	DC, OE	$WTP_{DC}=WTP_{OE}$
Brown et al. (1996)	Road removal in the North Rim of Grand Canyon	Public	DC, OE	$WTP_{DC}>WTP_{OE}$
Contingent Valuation Studies				
Cameron, Poe, Ethier, Schulze (1999)	Green Power Program	Public	DC, OE, PC	$WTP_{DC} >WTP_{PC}> WTP_{OE}$
Reaves, Kramer, and Holmes (1999)	Recovery of an endangered species	Public	DC,[5] OE, PC	$WTP_{DC}=WTP_{OE}= WTP_{PC}$
Ready, Navrud, and Dubourg Forthcoming	Avoidance of an episode of illness	Private	DC, PC[6]	$WTP_{DC} > WTP_{PC}$
Bohara et al. (1998)	Protection of Instream Flows	Public	DC, OE	WTP_{DC} WTP_{OE}[7]
Lunander (1998)	Preview of a movie	Private	DC, OE	$WTP_{DC} > WTP_{OE}$
Welsh and Poe (1998)	Reduced fluctuations in Glen Canyon Dam releases	Public	DC, PC	$WTP_{DC}>WTP_{PC}$
Loomis et al. (1997)	Art Print	Private	DC, OE	$WTP_{DC}=WTP_{OE}$

[4]The OE treatment involved a Vickery auction.

[5]The DC question used a double-bounded format where respondents who said yes to the initial offer amount were asked a follow-up question with a higher offer amount and respondents who said no to the initial offer amount were asked a follow-up question with a lower offer amount.

[6]Ready, Navrud and Dubourg refer to the payment card treatment as open-ended. However, the treatment was like a payment card in that respondents were shown a card with offer amounts and asked to check the amount they would pay.

[7] $WTP_{DC} > WTP_{OE}$ when a log normal distribution was used and $WTP_{DC} = WTP_{OE}$ when a weibull or gamma distribution was used.

Table 1: Recent Elicitation Studies since 1990

Authors	The Good	Public or Private Good	Response Formats	Results
Donaldson, Thomas, and Torgerson (1997)	A bone mineral density scan	Private	OE, PC	$WTP_{PC} > WTP_{OE}$
Boyle et al. (1996)	Ex post WTP to hunt moose in Maine	Private	DC, OE	$WTP_{DC} = WTP_{OE}$
	WTP of individuals who applied for a moose hunt permit but did not get one	Private	DC, OE	$WTP_{DC} > WTP_{OE}$
	Creation of a local response center to clean up oil spills	Public	DC, OE	$WTP_{DC} = WTP_{OE}$
Brown et al (1996)	Road Removal in the North Rim of the Grand Canyon	Public	DC, OE	$WTP_{DC} > WTP_{OE}$
Ready, Buzby, Hu (1996)	Food safety improvements	Private	DC, PC	$WTP_{DC} > WTP_{PC}$
Holmes and Kramer (1995)	Protection of a forest ecosystem	Public	DC, PC	$WTP_{DC} > WTP_{PC}$
McFadden (1994)	Wilderness Preservation	Public	DC, OE	$WTP_{DC} > WTP_{OE}$
Kriström (1993)	Protection of forest areas in Sweden	Public	DC, OE	$WTP_{DC} > WTP_{OE}$
Kealy and Turner (1993)	Candy Bar	Private	DC, OE	$WTP_{DC} = WTP_{OE}$
	Reduction in acid rain damage in Adirondacks	Public	DC, OE	$WTP_{DC} > WTP_{OE}$
Johnson, Bregenzer, and Shelby (1990)	Permit for one whitewater recreation trip on the Rogue River	Private	DC, OE	$WTP_{DC} > WTP_{OE}$

Figure 1: Willingness to Purchase Questions

Dichotomous Choice

Are you willing to purchase a [insert kw amount] kilowatt hour block of wind generated electricity each month for twelve months at an additional cost of $ [insert dollar amount] per month (that is $ [insert dollar amount] for the year)? ***(CIRCLE ONE NUMBER)***

1 No
2 Yes

Payment Card

MG&E is currently able to offer you wind generated power according to the terms listed in this table.

Size of block	Extra cost per month	Extra cost per year
50 kilowatt hours	$2	$24
100 kilowatt hours	$4	$48
200 kilowatt hours	$8	$96
300 kilowatt hours	$12	$144
500 kilowatt hours	$16	$192
600 kilowatt hours	$24	$288

Are you willing to purchase any of the blocks or wind generated electricity at the cost which are described in this table? ***(CIRCLE ONE NUMBER)***

1 No
2 Yes ---> Circle the particular kilowatt block you wish to purchase and fill out the form below:

50 100 200 300 400 500 600

Table 2: Response to Willingness to Participate Question

	Dichotomous Choice (N=648)	Payment Card (N=237)
Yes	24% (n=153)	37%(n=87)
No	76% (n=495)	63%(n=150)

Table 3: Response Distributions

	Dichotomous Choice (% Yes)	Payment Card (%)
$24	47%	33%
$48	35%	31%
$96	23%	9%
$144	18%	13%
$192	18%	7%
$240	16%	3%
$288	7%	3%

Figure 2: Empirical Survival Functions

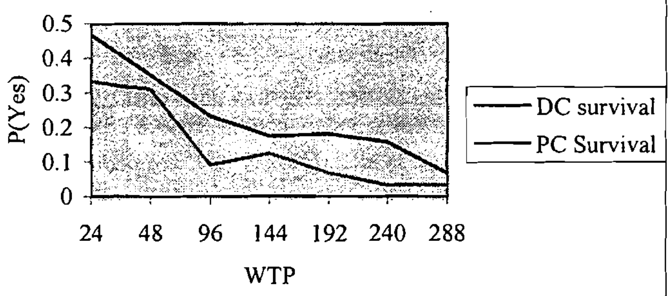


Table 4: Logit Models

	Dichotomous Choice (N=648)	Payment Card (N=249)	Pooled (N=897)
Constant	-0.173	-0.291	-0.138
	(0.163)	(0.142)	(0.099)
Offer Amount	-0.008	-.015	-0.010
	(0.001)	(0.002)	(0.001)
-2*Log Likelihood	664.59	617.20	1316.91
Hanemann's Estimate of Mean	$81	$36	$62

Table 5: Multivariate Models

Independent Variables	Coefficient (Standard Error)		
	Dichotomous Choice (n=518)	Payment Card (n=190)	Pooled (n=708)
Constant	-4.713[a] (.824)	-2.882[a] (1.050)	-4.073[a] (.635)
offer amount ($24, $48, $96, $144, $192, $240, $288)	-.015[a] (.001)	-.024[a] (.002)	-.017[a] (.001)
I can't afford to pay the extra cost of wind power (1=agree or strongly agree; 0=otherwise)	-2.400[a] (.390)	-2.782[a] (.589)	-2.531[a] (.323)
I think it is a good idea for MG&E to use wind power to meet the growing demand for electricity in the Madison area (1=agree or strongly agree; 0=otherwise)	1.201[a] (.444)	1.123[a] (.568)	1.110[a] (.337)
I like the idea of wind power (1=agree or strongly agree; 0=otherwise)	1.241[a] (.605)	1.185 (.913)	1.322[a] (.496)
I felt if I said I would buy the wind power, more renewable energy sources like wind, solar, and biomass are likely to be used in the future (1=agree or strongly agree; 0=otherwise)	1.966[a] (.334)	1.831[a] (.418)	1.792[a] (.247)
I would rather encourage conservation of electricity than pay to develop renewable energy sources (1=agree or strongly agree; 0=otherwise)	-.420 (.336)	-.926[a] (.426)	-.469 (.252)
I am willing to make personal sacrifices for the sake of the environment (1=agree or strongly agree; 0=otherwise)	1.318[a] (.446)	1.076[a] (.534)	1.199[a] (.337)
gender (1=male; 0=female)	-.490[a] (.248)	-.442 (.376)	-.477[a] (.203)
income	.000 (.000)	.000 (.000)	.000 (.000)
-2 log likelihood	837	381	1234

[a] significant at $\alpha = .05$ level
LR=27; $\chi^2_{\alpha=.05;9}=16.93$

References

Alberini, A. 1995. Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models. *Journal of Environmental Economics and Management* 28 (3):287-306.

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

Bohara, A.K., M. McKee, and R.P. Berrens. 1998. Effects of Total Cost and Group-Size Information on Willingness to Pay Responses: Open Ended vs. Dichotomous Choice. *Journal of Environmental Economics and Management* 35:142-163.

Boyle, K.J., F.R. Johnson, Daniel W. McCollum, W.H. Desvousges, R.W. Dunford, and S.P. Hudson. 1996. Valuing Public Goods: Discrete versus Continuous Contingent Valuation Responses. *Land Economics* 72 (3):381-96.

Brown, Thomas C., Patricia A. Champ, Richard C. Bishop, and Daniel W. McCollum. 1996. Which Response Format Reveals the Truth about Donations to a Public Good? *Land Economics* 72 (2):152-166.

Cadsby, Charles Bram, and Elizabeth Maynes. 1999. Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence. *Journal of Public Economics* 71:53-73.

Cameron, Trudy Ann, and Daniel D. Huppert. 1991. Referendum Contingent Valuation Estimates: Sensitivity to the Assignment of Offered Values. *Journal of the American Statistical Society* 86 (416):910-918.

Cameron, Trudy Ann, Gregory L. Poe, Robert G. Ethier, and William D. Schulze. 1999. Alternative Nonmarket Value-Elicitation Methods: Are the Underlying Preferences the Same? Unpublished Manuscript

Carson, Richard T., Theodore Groves, and Mark J. Machina. 1999. Incentive and Informational Properties of Preference Questions. Unpublished Manuscript.

Frykblom, P., and J.F. Shogren. 2000. An Experimental Testing of Anchoring Effects in Discrete Choice Questions. *Environmental and Resource Economics* 16:329-341.

Huang, Ju-Chin, and V. Kerry Smith. 1998. MonteCarlo Benchmarks for Discrete Response Valuation. *Land Economics* 74 (2):186-202.

Kealy, Mary Jo, and R.W. Turner. 1993. A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations. *American Journal of Agricultural Economics* 75 (2):321-331.

Loomis, John, Thomas C. Brown, Beatrice Lucero, and George Peterson. 1997. Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation. *Environmental and Resource Economics* 10:109-123.

Lunander, Anders. 1998. Inducing Incentives to Understate and to Overstate Willingness to Pay within the Open-Ended and the Dichotomous-Choice Elicitiation Formats: An Experimental Study. *Journal of Environmental Economics and Management* 35:88-102.

McFadden, D. 1994. Contingent Valuation and Social Choice. *American Journal of Agricultural Economics* 76 (November 1994):689-708.

Ready, Richard C., Stale Navrud, and Richard W. Dubourg. Forthcoming. How do Respondents with Uncertain Willilngness to Pay Answer Contingent Valuation Questions? *Land Economics*:26.

Reaves, Dixie Watts, Randall A. Kramer, and Thomas P. Holmes. 1999. Does Question Format Matter? Valuing an Endangered Species. *Environmental and Resource Economics* 14:365-383.

Rowe, R.D, William D. Schulze, and W.S. Breffle. 1996. A Test for Payment Card Biases. *Journal of Environmental Economics and Management* 31 (2):178-185.

Sheatsley, Paul B. 1983. Questionnaire Construction and Item Writing. In *Handbook of Survey Research*, edited by P. H. Rossi, J. D. Wright and A. B. Anderson. San Diego: Academic Press, Inc.

Welsh, Michael P., and Gregory L. Poe. 1998. Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach. *Journal of Envrionmental Economics and Management* 36:170-185.

An Experimental Economics Test of Factors Encouraging the Voluntary Provision of Public Goods

Donald J. Epp
Professor of Agricultural Economics
<dje2@psu.edu>

and

John D. Wicinas
Graduate Student
Department of Agricultural Economics and Rural Sociology
The Pennsylvania State University
University Park, PA 16802

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Partial financial support for this research came from the Department of Agricultural Economics and Rural Sociology, The Pennsylvania State University

Introduction

Public Goods have long been recognized as an example of what some call market failure. The non-rival and non-excludable characteristics of public goods make it nearly impossible for would-be producers and consumers of the good or service to interact in a market in such a way that the outcome is socially efficient (Hartwick and Olewiler). The characteristics of a public good make it possible for individual consumers to enjoy the benefits of consuming whatever amount of the good is produced without helping to pay the costs of producing the good. This is called free-riding. The likelihood of free-riding usually leads to the conclusion that some form of coercion is needed to require those who benefit from consuming the public good to pay something toward covering the costs of producing the good. In the usual case, government with its power to tax is chosen as the agent to provide the public good and to levy taxes to pay the costs.

Yet, there are notable examples of public goods that are provided without government intervention, though admittedly not at the socially optimal level. Any time the marginal benefit to an individual exceeds the marginal cost of producing one or more units of a public good, we can expect to see that amount produced. The providing individual reaps a benefit equal to or in excess of the costs of producing that amount of the public good and the remaining interested individuals may free-ride, consuming the good without payment.

In other cases we can observe individuals paying for the provision of goods and services consumed by others. In these cases of charitable giving it is assumed that the individuals receive some satisfaction from the act of giving or the improved welfare of the recipients. Charitable giving is often associated with providing private goods and servicesBfood, clothing or shelterBto those less financially fortunate than the contributor. But, charitable giving need not be restricted to providing private goods; provision of public goods is also supported by voluntary contribution. An example in the United State is the Public Broadcast System, which relies in significant part on voluntary contributions from individuals who, if they chose to, could watch whatever program was provided without contributing. Another example is found in contributions to organizations which provide protection for natural areas or particular species of wildlife. The continued existence of the protected objects is a public good. (A more detailed account of concepts associated with this topic may be found in Sugden (1982 and 1993) and in Cox.)

Research Question

The observation that public goods may be provided voluntarily by individuals in certain circumstances raises the question: What conditions encourage voluntary contribution to the provision of a public good? Answers to this question can be obtained in a variety of ways. We chose to conduct an economic experiment.

Hypotheses

1. Knowledge of the nature of public goods and the benefits which may accrue to the provision of greater amounts of a particular public good may encourage people to contribute voluntarily. Further, information about the contribution process, in this case the process of the experiment, is expected to encourage a greater level of voluntary contribution. We, therefore, hypothesize that increased levels of information and knowledge will lead to greater voluntary contribution to a public good.

2. Communication among participants can be a source of information as individuals share their knowledge of the provision of public goods. Communication among participants may also be used for psychological purposes to create an atmosphere conducive to giving or for one or more participants to build a feeling in others of an obligation to contribute. We hypothesize that increased levels of communication among the group will lead to greater voluntary contribution to a public good.
3. Many people have a strong sense of fair play. They resent cheating, especially if others gain advantage over them by not following the rules or agreements. Many people are more likely to contribute to the provision of a public good if they can be certain that others are also contributing. We hypothesize that providing assurance that others are contributing will increase the level of voluntary contributions to a public good.
4. Some of the highest levels of voluntary contribution to the welfare of others are observed among members of groups which provide mutual help and an extensive web of relationships among members. This includes certain religious groups whose members live in a community and have limited contact with others. In such communities, failure to conform to the group standard on one issue endangers all of one=s relationships in the community. A serious breach of expectations may result in having the rest of the community sever all ties with the violating individual. Exile and shunning are examples of such punishment. We hypothesize that groups with an extensive web of relationships will exhibit greater levels of voluntary contribution to a public good.

The Experiment

The economic experiment employed a game in which participants allocated points so as to increase their monetary reward from participating in the game. Hypotheses were tested by combinations of controls on the membership of the groups playing the game and the rules under which the game was played. We recruited three groups of undergraduate students at Penn State University. For each group we recruited nine participants, though in two cases fewer participants showed up for the experiment even though we had recruited extra members to allow for no-shows. Thus, we had one group of nine participants, one group of eight and one group of seven. We were able to rearrange the experimental setting to accommodate the smaller groups and maintain the essential elements of the experiment.

In recruiting participants, we promised that each person chosen to participate in the experiment would be guaranteed to receive $10 for about two hours of their time. They were told that their play of the game could results in a significantly higher reward, but they would receive at least $10 regardless of their play. Two of the groups were recruited from an undergraduate course that introduced students to environmental and resource economics (course groups). These students were promised a nominal (from the teacher=s viewpoint) number of extra credit points for the course in addition to the $10 guarantee. While these students were all from the same course, the course enrollment was large and most of the students in the course either did not know each other or had only a few common student organization memberships. (The experiment took place before public goods were discussed in the course.) The third group was recruited from a social fraternity at Penn State (fraternity group). The members of this group had a variety of academic interests and did not have any particular connection to environmental issues. A fraternity group was chosen to provide a group with a significant web of relationships outside of the experiment. A fraternity provides a number of communal decisions where individuals contribute, in some cases the contribution is required and the amount specified, while in other cases contribution is voluntary. Thus, the members of this group were also experienced in voluntary contribution to public goods, though it is doubtful that any of them would have used the economic terms to describe their experience.

We varied the rules of the game and the information we provided to test other characteristics hypothesized to determine levels of contribution,. The level of information about public goods and the advantages of contributing to the provision of public goods was included in a short hand-out and discussion with two of the groups. One of the course groups served as the control on informationBthey were provided only a brief introduction to how the game is played, but no information about public goods. The second course group was given instruction on how the game is played and information about how contributing to the public good can increase one=s reward. The information included three short examples where the participants were led through a calculation of the number of points they would receive under various levels of their contributions and those of other players. The fraternity group was given the same introduction to how the game is played as the other groups received and a short presentation reminding them of situations in fraternity life that are similar to public goods.

Variation in the levels of communication and the assurance that others contributed was provided by systematically varying the rules of play during the experiment for each group. The level of communication was expanded in three steps during the experiment and a method of assurance about the level of each participant=s contribution was introduced near the end of the experiment.

Rules of the Game

The experiment was presented as a game. Each group met at a different time and each group was seated around a U-shaped table with three participants on each of the three sides of the table. Each person was identified by a placard with a letter at his or her position. Letters were from A to I inclusive for the group with nine participants. The letter I was removed for the group with eight participants and the letters H and I were removed for the seven person group. Movable chairs were provided for the participants. This permitted them to shift easily from a spacing where each person could write without neighboring players being able to read what was written to a spacing where three players could converse without being heard by players in other sub-groups.

The play of the game consisted of three sets of 10 rounds in each set. That provided 30 rounds in total for each group. Each player received 25 points at the beginning of each set. The players were to decide in each round how many points to place in a Private Account and how many to place in a Group Account. All available points had to be allocated to an account. This allocation was indicated by writing the number of points to be contributed to each account on a form that was handed to a game marshal in each round. The points contributed to the Private Account were returned to the player at the end of the round. The points contributed to the Group Account by all of the players was totaled, multiplied by 1.5, and the resulting total divided equally among all players regardless of their level of contribution to the Group Account. Thus, the Group Account acted as a public good (non-rival and non-excludable) whose magnitude depended on voluntary contribution. The multiplication factor was chosen to provide a significant, but not overwhelming advantage to contribution to the Group Account. The amount of the Group Account payout for the round was added to each individual=s contribution record, added to the number of points contributed to the Private Account in that round, if any, and returned to the appropriate player. The total number of points at the end of each round was the number of points a player had available to use in the next round. Play continued for 10 rounds. At the end of each set (10 rounds) the total points for each player were recorded and a new set of 10 rounds began with each player receiving 25 points to begin play.

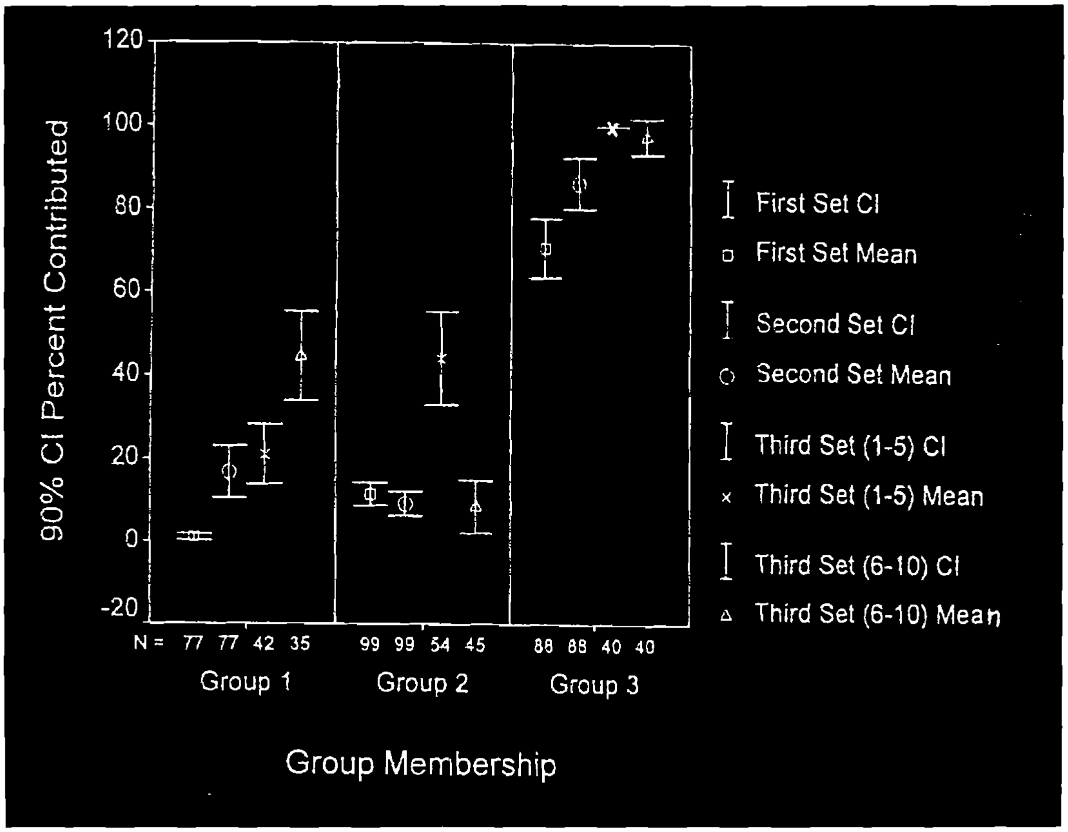
The first set of 10 rounds was conducted with no communication among players. Each individual acted alone in deciding his or her level of contribution to each account. In the second set of 10 rounds the players were divided into three sub-groups of three players each. When there were fewer than nine players, sub-groups consisted of two players each. Members of a sub-group (seated together on one side of the U-shaped table) could discuss strategy about contributing to the two funds, but they moved apart to record their own decision on their contribution form. This means that there was communication among a limited set of the players, but each individual=s contribution was confidential. Players had no assurance that other members of their group actually contributed according to any agreement they had reached in their discussion.

In the third set of 10 rounds, players were permitted to discuss strategy with all of the other players. Again, they marked their contribution forms in secret. Up to this point, players could not be certain what other individuals contributed to the Group Account. But, several of them were able to calculate what the pay-out from the Group Account would have been had all complied with the agreed level of contribution and to note that there was free-riding, even with communication among the entire group. At the completion of round 5 of the third set, assurance was introduced for each group. It was announced that for the remaining five rounds the contribution of each player would be posted where it could be seen by all players. In this way, players could determine if anyone did not follow any agreements reached during the discussion period.

During the entire experiment, players were asked to write their thoughts and observations on a sheet provided for that purpose at the end of each round. Most players recorded their thoughts for at least some of the thirty rounds that they played. These written comments often reveal that the individual knew the advantages to free-riding and several cases of frustration when a player realized that people were not cooperating or following agreements. At the end of the third set the points for each player were totaled and the players were paid one cent per point, with a minimum payment of $10.

Results of the Experiment

The results of the experiment are shown in Figure 1. (While the hypotheses are worded in a positive manner in this paper, the statistical analysis was conducted on null hypotheses in each case.) The most immediately obvious result is the much higher levels of contribution to the Group Account by the members of Group 3. Starting with the first set the mean contributions to the Group Account were statistically significantly higher than for either of the other two groups in any set. (Non-overlapping confidence intervals was used as the statistical test.) In the third set (rounds 1-5) every member of the group contributed 100 percent of his points to the Group Account. The slightly lower mean for rounds 6-10, with posting of the contributions of each



player, is explained by a single player=s error. In Set 2 one member placed his point allocation in the wrong placeBthe Private FundBin the ninth round. His commentary diary indicates that he realized his mistake and placed all of his points into the Group Account in the tenth (last) round of that set. But, the other players in Group 3 noticed the drop in points contributed to the Group

Account, assumed that others were starting to free-ride, and most players contributed all, or nearly all, of their points to the Private Account in the final round. In Set 3, one player remembered what had happened in Set 2 and determined to beat the other players to free-riding by placing all of his points in the Private Account in the ninth round. He announced this intention during the group discussion, but most players did not believe him. When he did as he announced and his allocation was posted, significant discussion erupted. His decision made it possible for him to end the game with substantially more points than the other players. It is probably sufficient to note that he was strenuously reminded of his obligations as a member of the fraternity as well as the human race. His remedy was to put all of his points into the Group Account in the final round, which restored everyone=s point total to what it would have been had he not been a free-rider in Round 9.

All of the outcomes with Group 3 compared to Groups 1 and 2 suggest that prior experience with voluntary contributions to a public good and a network of relationships among those in the group leads to significantly higher levels of contribution during the experiment. Tests of other hypotheses compare outcomes both within and across groups.

The role of information about public goods and the advantages of contributing to the Group Account in this experimental game is reflected in the difference in the mean contributions in Set 1 for the three groups. The Set 1 average contribution to the Group Account for Group 2 is significantly higher than for Group 1. Set 1 was played immediately after Group 2 received brief instruction about public goods. Group 1 received no information. The information given to Group 3 was to point out the public goods nature of some of their fraternity activities and the similarity of the Group Account to some of their fraternity activity funds. The mean contribution for the members of Group 3 in Set 1 is significantly higher than that of either Group 1 or Group 2. Whether this is due to an interaction of the information with experience or just the effect of experience cannot be determined with the available data.

If one examines the difference in mean contributions between Set 1 and Set 2 within a given group, we notice that Group 1 significantly increased their contributions in Set 2, although still not at a very high level. Comments made by some of the players indicate that even though they had some practice rounds before beginning Set 1, they really did not understand how the rules of the game could be used to their advantage until they had experimented a bit during Set 1. This type of learning may account for the insignificant difference between the average contribution of Group 1, Set 2 and Group 2, Set 1 (which followed instruction on public goods and the game).

The breadth of communication increased from set to set for each group. The increase in communication between Set 1 (no communication) and Set 2 (communication in sub-groups) was accompanied by increased contributions to the Group Account in Groups 1 and 3, but not in Group 2. The mean contribution by Group 2 was not significantly different between Sets 1 and 2 which seems to have been the result of players free-riding during Set 2. The free-riding was so extensive that the comments players wrote indicated either that they were taking advantage of other players (written by the free-riders) or that they knew they were being taken advantage of by other players and were going to give up and put everything in the Private Account (written by occasional contributors to the Group Account). By the time Group 2 reached Set 3, any sense of community among the players was gone and everyone was holding everything in the Private Account, even when the contributions of each player were posted in the last 5 rounds of Set 3. Group 1 did not exhibit an increase between Set 2 and the first half of Set 3, but the introduction of posted contributions at round 6 of Set 3 produced a significant increase in contributions to the Group Fund. Still, the average of those contributions was barely 50% of the total points available. Group 3 had significant increases in the level of contribution from Set 1 to Set 2 and from Set 2 to Set 3. The last five rounds of Set 3 were discussed above.

Posting each player=s contribution to the Group Fund had a significant positive effect on Group 1, no effect on Group 3 (they were at the 100% level of contribution already), and Group 2 experienced a significant drop in contribution level. The experiment developed in such a way that it did not really test the role of assurance, since Group 2 became extremely competitive and Group 3 was so cooperative that there was no room for assurance about contributions to increase.

Conclusions

Hypothesis 1Bthe role of information. We found a statistically significant, but small, effect where information increased contributions, at least initially.

Hypotheses 2Bthe role of communication. Broader communication gives greater levels of contribution to the Group Account. This is not a strong finding, as Group 2 showed that even communication among a broader group does not always produce greater cooperation. In fact, some of the player comments suggest that in Set 2, where players communicated within sub-groups, some players learned to free-ride from other members of their sub-group.

Hypothesis 3Bassurance that others are contributing the agreed amount will increase levels of contribution. The test of this hypothesis was confounded by other factors. Group 1 showed the expected increase. Contributions in Group 2 had fallen to nearly zero as free-riding was so frequent as to destroy any expectation of cooperation in contributing to the Group Account. Group 3 was already contributing 100% of their points to the Group Account before assurance was introduced.

Hypothesis 4Bexperience with public good provision and a network of relationships among group members will increase contributions to the Group Account. The experiment showed this to be the variable with the strongest impact on the level of Group Account contributions. The fraternity group contributed the highest percentage of points to the Group Account in each Set and was able to avoid the poisoning effects of free-riding, although individuals occasionally experimented with free-riding. Comments recorded by the players noted that, although they could determine that some players were free-riding or at least contributing less than they agreed, they knew from their fraternity experience that group cooperation was best for everyone. Once Group 3 reached a 100% contribution level, the members of the group were extremely vehement toward the individual who put all of his points into his Private Account in the next to last round of Set 3. The players did not take time to write down their reactions; they addressed them pointedly to the individual who had been identified in the posted contributions. Frequent reference was made by other players to the proper behavior toward one=s fraternity brothers. It was clear to an outsider, as well as to the players, that membership in the fraternity and the many different relationships they shared was supposed to induce individuals to behave for the good of the group. And those appeals and ties appear to have worked in the experiment.

Comment on Method

The structure of the experiment may have influenced at least one of the results. The monetary incentive to participate was provided in two parts. One provided players a reward (one cent per point) based on the outcome of their play in the game. The other was a guarantee that they would receive at least $10 for participating. The players in Group 2, and to a much lesser extent the players in Group 1, realized by the time they reached Set 3 that it was extremely unlikely that their payout at the end of the game would reach $10. Thus, they would receive the guaranteed amount regardless of how they played in the last set. This may have contributed to the high level of free-riding which produced little or no return since all players were trying the same strategy.

A more incentive compatible reward mechanism would have been to establish a fixed payment for participating and allowed the play of the game to determine how much extra compensation the players received. Since a given player=s payout depends heavily on the strategy and conduct of other players, the base reward must be large enough to appropriately compensate for the time spent in the experiment without obliterating the incentive to continue honest play of the game.

References

Cox, Donald. 1987. Motives for Private Income Transfers. J. of Pol.Econ. 95:508-546.

Hartwick, John M. and Nancy D. Olewiler. 1986. The Economics of Natural Resource Use. Harper and Row, Publishers, Inc. New York.
Sugden, Robert. 1982. On The Economics of Philanthropy. The Econ. J. 92:341-350.
Sugden, Robert. 1993. Thinking as a Team: Towards an Explanation of Non-selfish Behavior. Altruism. 69-89.

Testing a Computer-Assisted Valuation Panel Approach For Valuing Watershed Ecosystem Restoration

Eric Huszar
U.S.D.A. APHIS

John Bergstrom
University of Georgia

Tom Holmes
U.S.D.A. Forest Service

Susan Kask
Western Carolina University

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Introduction

The Little Tennessee River watershed is located in Georgia, North Carolina, and Tennessee Figure 1. The watershed encompasses 10,783 acres, including 18 rivers and streams and 26 lakes (EPA website). The Little Tennessee River (LTR) originates in Rabun County Georgia; it flows north into North Carolina before terminating at Fontana Dam, just south of the Smokey Mountains.

Historically the Little Tennessee River watershed has been known for its logging, agriculture and mining industries, however recently tourism, recreation and the draw of living in an aesthetically pleasing environment has led to a tremendous increase in the population of people who visit and live within the watershed. In the last twenty years the population has doubled leading to concerns about the future health of the watershed and the ecosystem services the watershed provides. The majority of land within the watershed is privately owned and private land use decisions have a major impact on ecosystem structure and function. For example, agricultural activities (such as watering cattle in streams) and development (housing and commercial development along the streams and creeks) influence water quality, a key parameter of ecosystem health.

Valuing ecosystem services is controversial because of the potential importance such values may have in influencing public opinions and policy decisions (Costanza, et al. 1998). Very little is also known about how the public values ecological goods and services produced by mixed public and private land ownership at a landscape scale. However a failure to quantify or value these ecosystem services could imply a zero value placed on them while Daley (1997) notes that ecosystem services in most cases have values larger than zero.

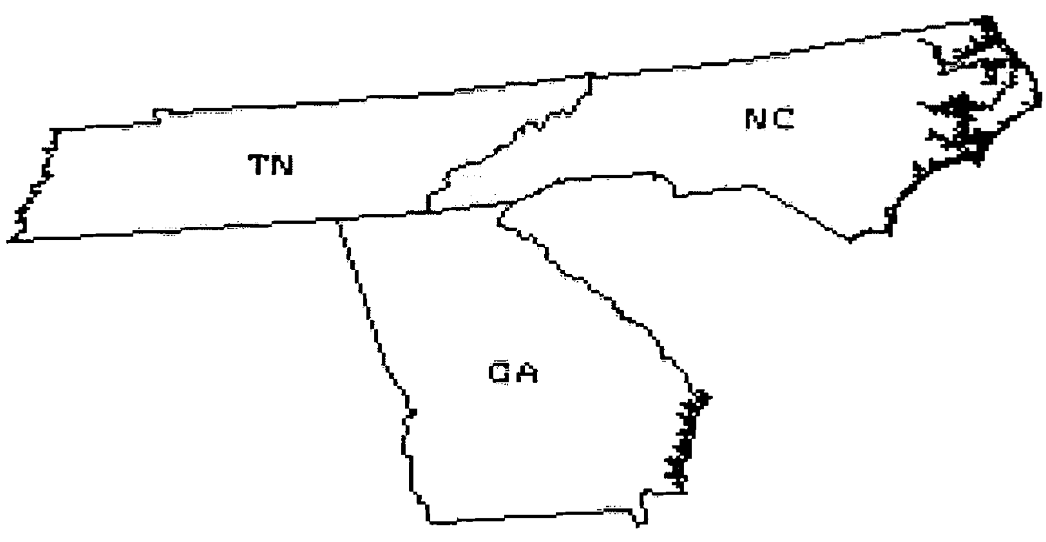


Figure 1: Little Tennessee River Watershed

Study Area (LTRW)

The objectives of this study were to develop and test general methodology for valuing ecosystem services and to identify and value particular ecosystem services present in the Little Tennessee River watershed. An in-depth review of the literature was conducted looking at past studies that could be applicable to ecosystem valuation and changing ecosystem characteristics. This includes but is not limited to stated preference studies addressing embedding effects, scope effects, substitutability, complementarity, and studies evaluating system- level properties such as biological diversity. With the help of focus groups an appropriate survey would be developed to evaluate changes in ecosystem characteristics, a site to implement the survey would be selected along with the presentation format, either through the mail, telephone, or in person, and appropriate quantitative methods would be used to evaluate the data collected through the survey implementation.

Methodology

To place a value on ecosystem services, a stated preference or contingent valuation (CV) survey instrument was designed based on the literature. Contingent valuation is a standardized and widely used survey method for estimating willingness to pay (WTP) or willingness to accept compensation (WTA) for use, existence, and bequest values for resources (Loomis 1996). While several methods can be used to value "use values" such as the travel cost model (TC) or the hedonic property method only a few can estimate the existence or bequest values "non-use values" associated with an ecosystem. CV surveys can reveal an existence value or what a person would pay just to know that a certain species, such as a trout is present in its natural habitat. Bequest values can also be revealed as the amount a person would be willing to pay just to know that future generations will have certain species such as trout present in there natural habitat.

CVM involves construction of a hypothetical market or referendum scenario in a survey. The proposed increase in the quantity or quality of the resource is communicated to the respondent in words and visual aids. The respondent uses the hypothetical market to state his or her support for or against a program along with their values associated with the program. The "Blue Ribbon" panel, which consisted of Kenneth Arrow, Robert Solow, Paul Portney, Edward Leamer, Roy Radner and Howard Schuman, concluded that CVM produces estimates reliable enough to be the starting point for a judicial or administrative determination of natural resources values (Arrow et al. 1993).

The concern with CV is the reliability and validity of the responses. Would these individuals really pay the amount stated in the survey? All the published studies to date have shown CVM derived responses of WTP for both use and non-use values to be reliable in test-retest studies (Loomis, 1989; Carson, et al., 1997). The NOAA panel has suggested the use of the referendum CV format due to its close proximity with market goods and voting situation. The respondent is given a discrete question: Here is an item that costs $X; will you "take it" or leave it? The referendum format closely represents the real world, for example a box of cereal costs $X you can either purchase it or leave it.

It is believed that the use of the referendum CV increases accuracy and validity in estimating the consumer surplus for environmental improvements.

This study hopes to avoid some of the shortcomings mentioned earlier such as sub-addivity effect WTP for changes in a collection of public goods is less than the sum of WTP for changes valued separately. The sub-addivity effect particularly relevant for ecosystem valuation. Sub-addivity can be avoided my using a comprehensive value for a set of ecosystem services and thereby reducing the possibility of double counting and the summation problem noted by Hoehn and Randall 1989. As noted from Loomis (2000) a baseline or current conditions in the watershed were described, allowing the respondent then to compare baseline conditions to improved or degraded conditions in the watershed.

Relevant Literature

Of the 3.5 million miles of rivers in the United States only 56% of them fully support the multiple use mandate adopted by the U.S. Forest in 1992. The multiple use mandate outlines that streams and rivers should supply drinking water, fish and wildlife habitat, recreation, and agriculture, as well as flood prevention and erosion control (USDA 1998). This multiple use mandate has lead to the increased interest in restoring rivers and streams so that they may once again support these multiple uses or ecosystem services.

Costanza et al completed one of the must ambitious and controversial studies dealing with valuing the world's ecosystem services (Costanza 1997). The authors estimated the economic value of 17 ecosystem services for 16 biomes based on published studies and existing data at $16-54 trillion per year. At the time of the study the global gross national product was estimated at $18 trillion per year. This study of the worlds ecosystem services valuation has lead to various heated discussion in the literature regarding the validity and importance of the study. Many of the issues surrounding the Costanza paper are discussed in a special section of *Ecological Economics 1998.*

Loomis et al. (1999) looked at the values of restoring certain ecosystem services along a 45-mile stretch of a river in Northern Colorado. A dichotomous choice willingness to pay question regarding purchasing increased ecosystem services such as dilution of wastewater, natural purification of water, erosion control, habitat for fish and wildlife, and recreation was administered to 100 participants. An interdisciplinary team of economists and scientist worked to develop management actions necessary to increase the level of ecosystem services. These management actions included a ten mile wide conservation easement along 45 miles of the river, restoring native vegetation along the river in the form of buffer strips and eliminating cropland and cattle grazing in these buffer strips, and reducing the amount of water diverted from the river from 50% to 70% to increase stream flows. A mean annual household willingness to pay of $252 was estimated for the increase in ecosystem service on the 45-mile stretch of the river.

Survey Design and Implementation

In order to collect data, one must first design a survey that includes detailed descriptions of the resources being valued. It is crucial then that the design of the CV survey clearly describe the goods to be valued. An initial step of designing a CV survey is to become familiar with the latest scientific research on the area of focus. This lead to a group of three economists and several scientist in the area meeting and talking about how the Little Tennessee River was being impacted from agricultural uses and development within the watershed.

The economists met with ecologist at the U.S. Forest Service Coweeta Research Laboratory to discuss the different possible ecosystem services and attributes of the river that should be valued. It was also learned that the river was already undergoing certain stream bank restoration measures aimed at improving water quality along the river. The Little Tennessee River association along with the Macon Soil and Water Conservation District was in charge of these stream bank restoration measures along the river. The two organizations informed the group of economist of the potential improvements in ecosystem services due to the stream bank restoration measures.

Management action plans in the Little Tennessee River included stream bank restoration along the Little Tennessee River and buffer zones along small streams and creeks in the watershed. Three different forms of stream bank restoration would be implemented to improve ecosystem services on the Little Tennessee River along with the total improvement of the watershed.

The first type of restoration measure consisted of fencing along the river to eliminate livestock from the river. A fence would be installed on either side of the river approximately 30ft from the stream bank. Fencing is established to keep livestock from using the stream for drinking water and thus damaging the stream bank, fencing also prevents livestock waste from directly entering the river. Farmers are then provided with funds to establish an alternative water source for livestock (such as watering tanks).

The second type of stream bank restoration dealt with rebuilding eroding stream banks with revetment. Revetment consists of large tree branches being attached along the streams bank. Revetment is used to stabilize the eroding bank, and rebuilt the integrity of it. The last type of stream bank restoration dealt with buffer zones along the river. Buffer zones provide a natural barrier between human activity (e.g. agricultural, residential development) and the river. This thus reduced the amount of pollutants that enter the river.

After becoming familiar with the best scientific knowledge available about the watershed, it was then important to test certain watershed programs aimed at improving ecosystem health. Five different focus groups were held so that illustration/pictures and text about these three forms of stream bank restoration could be represented to the public, and we could receive feedback on them.

The first focus group was held at the University of Georgia on April 24, 2000. This initial focus group was held primarily as a warm up to the focus groups that would be held around the sampling area of Macon County in North Carolina. A PowerPoint presentation representing maps and pictures of the area was very helpful to the group. The maps and pictures allowed the group to free associate with thoughts, feelings, and knowledge they had about the Little Tennessee River. While this audience was geographically biased from are sampling area, they showed a large degree of concern for the river and the restoration efforts taking place on it. With these helpful suggestions and comments received it was time to present the material to a group of people in the study area.

The second focus group was held at the Nantahala Power and light building in Macon County North Carolina on May 18, 2000. Different civic organizations were asked to attend the focus group with an incentive of $25 to each person who attended. This group of participants was much more familiar with the Little Tennessee River than the initial focus group participants in Georgia. We found that certain terminology used to describe restoration was not clear, and a further explanation of terms like revetment was needed. The group also revealed some characteristics/attributes of the LTR that were important to them in restoring the river. Attributes that came to mind before restoration were; erosion/sediment loss, poor quality (fish, wildlife), ugly, river pollution, stagnant water, and dead trees compared to after restoration; return of wildlife, erosion and sediment control, recreation tourism, clear water, improved quality, pretty, pride, and happy farmers.

The third focus group was also held at the Nantahala Power and light building in Macon County North Carolina on June 27, 2000. Different civic organizations were asked to attend the focus group with an incentive of $25 to each person who attended. This focus group was interesting in the fact that we had some participants return from the previous focus group to give use feedback on how well we had incorporated there ideas and suggestions into the illustration and pictures of the LTR. The different watershed management plans were introduced along with some type of payment vehicle.

The fourth focus group was held in Cullowhee North Carolina at Western Carolina University on August 14, 2000. The participants again received $25 each for participating. The information from the previous groups was now in a survey format describing different watershed management plans along with a valuation aspect. The material was presented to the group using interactive computers. The computer allowed the participants to learn about the river and the different management plans as well as starting to think about valuing these actions.

The Fifth and final focus group was again held in Cullowhee North Carolina at Western Carolina University on September 22, 2000. The participants received $25 dollars for attending. This last focus group was the final test to make sure that the survey was clear about the different watershed management programs to be valued. It was also are last chance to make sure that everyone was confident and comfortable enough to take the survey over a computer.

By using this computer based survey implementation (Ci3 software) we could both inform the survey participant about the three different stream bank restoration measures as well as ask them questions at the same time. The problem with using a written CV survey is the more complex valuations require a pamphlet to describe and illustrate the different hypothetical scenarios in addition to the questionnaire. So incorporating both information about the Little Tennessee River and questions about the watershed seemed to simplify the whole process compared to presenting informational material followed with the questionnaire as would need to be done in a written survey.

The computer based CV survey described four different watershed management programs that would improve/affect ecosystem services such as erosion control, purification of water, fish and wildlife habitat, and preservation of natural areas. As mentioned above these management programs involved different levels of stream bank restoration along the Little Tennessee River along with buffer zones along other small creeks and streams.

Five different indicators of ecosystem services were identified with the help of scientist and the focus groups; 1) Game Fish including species like trout, bass, and other sport fish, 2) Water Clarity or the amount of suspended sediment in the water, 3) Wildlife habit in buffer zones including deer, birds and other animals, 4) Allowable Water uses including safe for boating, fishing, or drinking, and 5) Index of ecosystem "naturalness" or how closely the area has been returned to its natural state. After describing and identifying these five indicators of ecosystem services a panel of respondents was presented with the different management programs and asked to evaluate them independently.

A valuation panel was used due to complexity of the different watershed programs to be described. Valuation panel data allows the survey to be set up in a given location within the sampling area, and provides the means for describing the programs in great detail with pictures and illustration of the management programs to be valued. Without a panel format, some other deliver method would need to be used such as in-person interviews, which can be very expensive and time consuming. In person interviews are also subject to potential interviewer bias.

Civic groups and other organizations in the Macon County area were contacted and asked if they would participate in the survey. Each individual who attended the survey would be paid $40 for there time. The survey panels were held in computer labs at Franklin high school and Southwestern College. A total of 191 respondents 18 years of age or older completed the survey. Table 1 shows how many people attended from each of the ten different organizations in Macon County.

Table 1 Civic Group Panels

Organization	**Number (191)**	**Percentage**
Cullasaja Elementary School	36	19%
Cowee Elementary School	32	17%
Franklin High School	11	6%
Union Elementary School	28	14%
Cartoogechaye Elementary School	22	11%
St. Francis	4	2%
St. Agnes Episcopal Church	7	4%
Franklin Lions Club	17	9%
Franklin Fire and Rescue	25	13%
Southwestern Students	9	5%

Approximately 70% of the sample consists of different school organizations; these organizations are made up of teachers along with parents whose children attend the school.

Preliminary Results

Sociodemographic statistics describing the 191 respondents sampled are presented in Table 2.

Respondents' familiarity with the LTR and water management issues are summarized in Table 2a. Many of the 191 respondents surveyed in Macon County were familiar with the location of the LTR as would be expected since the river is located in Macon County. Not only were respondents familiar with the location of the river but many respondents approximately 63%, were aware that restoration measure had taken place along the river. A majority of respondents seemed to participant in some sort of activities involving the river be it, fishing/hunting or just simply enjoying a drive along the river for viewing pleasure.

Table 2 Descriptive Statistics for N=191				
Variable	Mean	Standard Error	Minimum	Maximum
AGE	45.86	12.17	21	77
GENDER 1 = male, 0 = female	0.387	0.488	0	1
INCOME	$53,115	$28,829	$10,000	$165,000
HOUSEHOLD SIZE	3.1	1.44	1	11
EDUCATION 1= elementary to 5 = Graduate Prof.	4.06	0.72	2	5
LIVE BESIDE LTR 1 = yes, 0 = no	0.523	0.5	0	1

Notes: The frequency for males and females was 117 females (61%) and 74 males (39%)
Education: 1= Elementary, 2 = JR. High, 3 = High school, 4 = College or Technical, and 5 = Graduate or Professional school. LTR = Little Tennessee River.

Table 2a Descriptive Statistics for N=191					
Variable	Description	Mean	Standard Error	Minimum	Maximum
FLTR	Familiarity with the Location of the Little Tennessee River, 1 = very familiar, to 4 = not at all familiar	1.59	0.767	1	4
ACT	During past 12 months did you participate in Fishing, walking hiking, hunting, wildlife viewing, Boating/rafting, or scenic drive to see the river 1 = yes, 0 = no	0.586	0.493	0	1
USES	In your opinion, which of the following uses in the LTR is most important 1 = Agriculture 2 = Residential, 3 = Recreation, 4 = Fish/Wildlife	2.82	1.2	1	4
AW	Before taking this survey were you aware of Restoration taking place along the LTR 1 = yes, 0 = no	0.633	0.483	0	1
RES	Familiarity with revetment, fencing, buffer zones 1 = very familiar, to 4 = not at all familiar	2.67	1.04	1	4

After being introduced to three different types of stream bank restoration measures; 1) revetment (tree limbs along the stream bank), fencing (keep cattle from the river), and buffer zones (natural areas at least 30 feet on each side of the river) in the form of pictures/illustrations and text, respondents were asked to rate their importance on a scale from 1 being not important to 7 being very important. Table 3 gives descriptive statistics for the entire 191 respondents sampled while Table 4 and Table 5 give descriptive statistics depending upon the version of the survey along with the management action taken. A description of the management plan is given along with the indicators of ecosystem services impacted.

Table 3 reveals that many of the indicators of ecosystem health along the LTR were very important. They felt restoration, fish habitat, wildlife habitat, erosion control, purity of water, and ecosystem naturalness should be protected. Out of a possible 7 being very important the mean score was 5.6. Tables 4 and 5 present information on the desirability of different management programs in relationship to the quality of attributes in LTR, such as the level of water clarity, level of allowable water uses, (drinking, swimming, etc) level of game fish, level of wildlife habitat, and level of ecosystem naturalness. These different levels of attributes ranged from low, moderate to high. Again they are asked on a scale from 1 to 7 how desirable the management program is to them. The programs with the highest quality of attributes were more desirable to the respondents.

Table 3 Descriptive Statistics for Scaling Questions (N = 191)

			Standard	Frequency *(1 = not important, 7 = very important)*						
Variable	**Description**	**Mean**	**Error**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
RES	Importance to establish revetments, fences, buffer zones along the river	5.66	1.4	2	4	8	31	20	58	68
BUF	Importance to establish buffer strips	5.56	1.41	3	5	3	34	35	47	64
FISH	Importance of fish habitat	5.74	1.32	1	2	3	38	30	38	79
WILD	Importance of wildlife Habitat	5.76	1.23	1	1	2	36	28	52	71
ERO	Importance of erosion control	5.88	1.22	1	2	0	30	29	49	80
WATER	Importance of water purity	6.05	1.17	1	1	1	24	24	46	94
ECO	Importance of ecosystem naturalness	5.6	1.3	1	2	7	38	25	59	59

Table 4 Desirability of different programs N = 95 (version 1)

Variable	Description	Mean	Standard Error	Frequency *(1 = not important, 7 = very important)* 1	2	3	4	5	6	7
CUR	How desirable is the Current management program (low water clarity, low level of uses, low level of game fish, low level of wildlife habitat, low level of ecosystem naturalness	2.77	2.12	42	16	5	11	5	6	10
ALT 1	How desirable is management alternative one (mod. water clarity, low level of uses, low level of game fish, a moderate level of wildlife habitat, low level of ecosystem naturalness	3.06	1.76	22	24	9	20	12	2	6
ALT 2	How desirable is management alternative two (mod. water clarity mod, level of uses, low level of game fish, mod. Level of wildlife habitat, mod. level of ecosystem naturalness	4.08	1.77	6	12	17	27	10	12	11
ALT 3	How desirable is management alternative three (mod. water clarity high level of uses, mod. Level of game fish, high level wildlife habitat, mod. Level of ecosystem naturalness	5.2	1.57	3	4	3	22	16	23	24
ALT 4	How desirable is management alternative four (low water clarity mod. level of uses, low level of game fish, mod. level wildlife habitat, low level of ecosystem naturalness	2.98	1.71	24	17	20	19	5	5	5

ble 5 Desirability of different programs N = 96 (version 2)

riable	Description	Mean	Standard Error	Fre 1	Fre 2	Fre 3	Fre 4	Fre 5	Fre 6	Fre 7
				(1 = not desirable, 7 = very desirable)						
JR	How desirable is the Current management program (low water clarity, low level of uses, low level of game fish, low level of wildlife habitat, low level of ecosystem naturalness	2.75	1.77	31	24	9	16	6	6	4
,T 1	How desirable is management alternative one (low water clarity, mod level of uses, low level of game fish, a mod. level of wildlife habitat, low level of ecosystem naturalness	3.07	1.64	16	29	14	20	7	6	4
,T 2	How desirable is management alternative two (mod. water clarity mod, level of uses, low level of game fish, mod. level of wildlife habitat, mod. level of ecosystem naturalness	4.01	1.35	4	10	14	37	18	10	3
,T 3	How desirable is management alternative three (mod. water clarity mod. level of uses, low. level of game fish, high level of wildlife habitat, high. level of ecosystem naturalness	4.82	1.33	1	3	8	33	16	25	10
,T 4	How desirable is management alternative four (high water clarity high level of uses, high level of game fish, high level of wildlife habitat, high level of ecosystem naturalness	5.75	1.4	1	4	3	12	9	28	39

Valuation Questions

Two different versions of the survey were developed to deal with different watershed valuation issues. The first version of the survey dealt specifically with restoring from 2 to 6 miles of stream bank along the LTR. The second version of the survey was very similar except another component was added to see how respondents felt about all small streams and creeks within the LTR watershed being protected with buffer strips. These two different versions of the survey are shown in Table 6 and Table 7. A sample of the valuation question asked is shown below.

Sample Question

The general questioning format for the valuation questions was for the respondent to consider that a vote is being held today in Macon County to approve or reject a management program for the Little Tennessee River watershed. The Management program would be one of the alternative stream bank restoration programs for the Little Tennessee River described in the previous sections. If you agree to support the program, your payment would be collected through an increase in the local sales tax you pay. The program will be implemented only if a majority of Macon County residents vote in favor of it. Please consider your current expenses before answering the following questions.

If a local county sales tax were to reduce your annual household income by $X each year for the next 10 years to support program X, would you vote in favor of it. (Taking into account your other expenses and the fact that you probably already pay something for water)

Table 6 Overview of Little Tennessee River Watershed Programs

Version 1	Current Situation	Program 1	Program 2	Program 3	Program 4
Indicator of Ecosystem service	No small streams protected by buffer strips + no new river restoration	No small streams protected by buffer strips + 2 miles of new river restoration	No small streams protected by buffer strips + 4 miles of new river restoration	No small streams protected by buffer strips + 6 miles of new river restoration	All small streams protected by buffer strips + no new river restoration
Game Fish	Low	Low	Low	Moderate	Low
Water Clarity	Low	Moderate	Moderate	Moderate	Low
Wildlife habitat in buffer zones	Low	Moderate	Moderate	High	Moderate
Allowable Water uses	Low	Low	Moderate	High	Moderate
Index of ecosystem "naturalness"	Low	Low	Moderate	Moderate	Low

Table 7 Overview of Little Tennessee River Watershed Programs

Version 2	Current Situation	Program 1	Program 2	Program 3	Program 4
Indicator of Ecosystem service	No small streams protected by buffer strips + no new river restoration	All small streams protected by buffer strips + no new river restoration	All small streams protected by buffer strips + 2 miles of new river restoration	All small streams protected by buffer strips + 4 miles of new river restoration	All small streams protected by buffer strips + 6 miles of new river restoration
Game Fish	Low	Low	Low	Low	High
Water Clarity	Low	Low	Moderate	Moderate	High
Wildlife habitat in buffer zones	Low	Moderate	Moderate	High	High
Allowable Water uses	Low	Moderate	Moderate	Moderate	High
Index of ecosystem "naturalness"	Low	Low	Moderate	High	High

Preliminary Model

For each of the different programs shown above in Tables 6 and 7, respondents were asked if they would vote to support the LTR protection program at the stated price or cost. Using the logistic model, we can statistically analyze a respondent's "yes" or "no" response. A binary response is recorded with a "yes" response indicated by Y=1 and a "no" response indicated by Y=0. If a "yes" is recorded for X$ then we know a particular person would pay at least X$, however he or she may be WTP more than X$. We can say that a "yes" response bounds the true WTP from below the bid offered. Similarly, in theory, then a "no" response will bound the true WTP from above the bid offered.

The logistic model can then be written as: Green (1997)

$$PROB(Y=1) = \frac{e^{\beta' x}}{1+e^{\beta' x}}$$

where, X = the characteristics that the sample population possesses.

Each of the eight different programs has been evaluated using logistic modeling. Tables 8 and 9 report the parameter estimates for each model. Table 8 present results for the first version of the survey dealing mainly with stream bank restoration. Table 9 presents results for the second version of the survey, which deals with stream bank restoration along with the idea of establishing additional buffer zones. After presenting these tables a brief discussion of the signs and significance levels of the explanatory variables is given.

While these are just some preliminary results, the explanatory variables that seem to be significant and positive are education, participate, and Gender. More educated respondents are more likely to contribute to the watershed management programs. It is also somewhat evident that people who participate in activities along the LTR such as hiking, wildlife viewing, and boating are more likely to support these watershed management programs. Gender seems to also be somewhat significant and positive.

The estimated bid coefficients for version two have an expected negative sign. As the bid amount is raised, we would expect that most people would become less likely to support the proposed watershed program. The bid coefficient in version two is not only negative but is highly significant in the first two programs. In version one however the bid coefficient is positive and only significant in two of the programs at the .10 level. This is troublesome and raised some concern on why people would be more likely to support the proposed watershed program as the bid amount is increased. Positive correlation with increasing quality of the LTR watershed over the alternative programs may be part of the explanation.

Table 8 Referendum Models for Survey Version One

	Program 1 N = 95	Program 2 N = 95	Program 3 N = 95	Program 4 N = 95
Variables				
Constant	-5.46 (2.07)***	-3.94 (1.88)**	-2.71 (1.69)*	-1.71 (1.72)
Aware restoration 1 = yes 0 = no	0.309 (0.49)	0.4 (0.47)	-0.17 (0.46)	-0.41 (0.47)
Participate 1 = yes 0 = no	0.26 (0.48)	0.26 (0.47)	0.60 (0.47)*	0.75 (0.47)*
House size	0.17 (0.20)	0.17 (0.18)	0.04 (0.18)	-0.14 (0.18)
Education	0.79 (0.32)**	0.79 (0.32)**	0.46 (0.29)*	0.39 (0.31)*
Sex 1 = male, 0 = female	0.61 (0.50)*	0.61 (0.47)*	0.35 (0.44)	0.73 (0.45)*
Age	0.01 (0.01)	0.01 (0.01)	0.005 (0.01)	-0.01 (0.01)
Bid amount	0.01 (0.007)*	0.01 (0.003)*	-0.001 (0.001)	-0.001 (0.007)
Log Likelihood	-58.32	-60.13	-62.63	-58.53
Probability of Yes At bid = $50	0.54	0.43	0.45	0.35
Pseudo R-squared	0.08	0.049	0.039	0.06

1) *White's standard errors, robust to misspecification, are in parentheses below the parameter estimate. (.01 = ***, .05 = **, and .10 = *)
2) Krinsky-Robb method is used for 95% confidence interval (1,000 draws)

Table 9. Referendum Models for Survey Version Two

	Program 1 N = 96	Program 2 N = 96	Program 3 N = 96	Program 4 N = 96
Variables				
Constant	-4.63 (1.66)***	-2.19 (1.59)*	-5.87 (2.01)***	0.65 (1.36)
Aware restoration 1 = yes 0 = no	0.24 (0.49)	-0.1 (0.49)	-0.19 (0.49)	0.27 (0.44)
Partcipate 1 = yes 0 = no	0.31 (0.50)	0.27 (0.48)	1.03 (0.54)*	0.44 (0.43)
House size	-0.16 (0.16)	0.17 (0.16)	-0.03 (0.14)	-0.08 (0.14)
Education	0.99 (0.35)***	0.65 (0.36)*	0.66 (0.39)*	-0.01 (0.22)
Sex 1 = male, 0 = female	0.30 (0.49)*	0.04 (0.46)	-0.13 (0.50)	-0.20 (0.45)
Age	0.02 (0.02)	0.001 (0.02)	0.03 (0.02)*	-0.01 (0.01)
Bid amount	-0.02 (0.008)**	-0.02 (0.008)***	-0.0016 (0.006)	-0.004 (0.003)*
Log Likelihood	-54.32	-54.13	-50.63	-64.44
Probablily of Yes at bid = $50	0.30	0.25	0.25	0.54
Pseudo R-squared	0.12	0.047	0.11	0.01

1) *White's standard errors, robust to misspecification, are in parantheses below the parameter estimate. (.01 = ***, .05 = **, and .10 = *)
2) Krinsky-Robb method is used for 95% confidence interval (1,000 draws)

Summary and Conclusions

For this research project, a CV survey to assess the economic values of stream bank restoration on the Little Tennessee River was developed with the input of several scientist and focus groups. It was evident in the focus groups that people seemed to be aware of the LTR and the stream bank restoration-taking place along it. The survey was administered through a computer (Ci3 software) using central facility valuation panel. The computer driven survey automatically recorded the respondents' answers for later data analysis. The computer survey reduced the time usually associated with making paper copies of the survey and then going back and entering all the data collected from the paper survey. Overall, the computer-assisted valuation panel approach appears to a promising approach for valuing complex environmental commodities such as watershed ecosystems.

The computer survey did have some drawbacks however; it was very time consuming to enter all the survey material in the appropriate Ci3 computer language. The Ci3 software was also less than user friendly and required some time talking with the producers of the software. The computer survey also needed to be placed on computers with an appropriate graphics package or else the pictures and illustrations were grainy and hard to see. This required the survey group to located computers within the sampling community that could be used for the central facility valuation panel.

A unique biding design was used for this survey and the current single bounded logit model applied to each of the 8 different programs does not sufficiently represent this design. The respondent is asked to value these programs independently, but in a sense this is a difficult task because the bid amounts for each of the different programs are linked together. Some type of lag model that could estimate these models together would be more appropriate than the current process used to value each of the programs independently from one another. Further estimation work along these lines by the authors is in progress.

References

Arrow, K.; R. Solow; E. Leamer; P. Portney; R. Radner; and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." Appendix I. Fed. Reg. 58: 4002-13

Costanza, R.; R. Arge; R. de-Groot; S. Farber; M. Grasso; B. Hannon; K. Limburg; S. Naeem; R. Oneil; J. Paruelo; R. Raskin; P. Sutton; and J. van den Belt; 1998. The value of the worlds ecosystem services and natural capital. Ecological Economics. Vol 25: 3-15.

Costanza, R.; R. Arge; R. de-Groot; S. Farber; M. Grasso; B. Hannon; K. Limburg; S. Naeem; R. Oneil; J. Paruelo; R. Raskin; P. Sutton; and J. van den Belt; 1998. The value of ecosystem services; putting the issues in perspective. *Ecological Economics*. Vol. 25: 67-72.

Dailey, G. 1997. Nature's Services: Societal Dependence on Natural Ecosystem. Island Press, Cavelo, Ca.

EPA Website 2000. www.epa.gov/surf2/hucs/06010202. Watershed info.

Hoehn, J., A. Randall. 1989. Too many proposals pass the benefit-cost test. *American Economic Review*. Vol 79: 544-551.

Loomis, B.J.; P. Kent; L. Strange; K. Fausch; and A. Covich 2000. Measuring the total economic value of restoring ecosystem services in an impaired river basin: results from a contingent valuation survey. *Ecological Economics* Vol 33: 103-117.

Loomis, B.J. 1996. Measuring the economic benefits of removing dams and restoring the Elwha River: Results of a Contingent valuation survey. *Water Resources Research*. Vol. 32: 441-447.

Loomis, B.J. 1989. Test-retest reliability of the contingent valuation method: a comparison of general population and visitor responses. *American Journal of Agricultural Economics*. Vol 71: 76-84.

United States Department of Agriculture. 1998. Stream Corridor Restoration, Principles, process, and practices by the Federal Interagency Stream Corridor Restoration Working Group. *(hhp://www.usda.gov/stream_restoration).*

Contingent Valuation Willingness to Pay with Respect to Geographically Nested Samples: Case Study of Alaskan Steller Sea Lion

Branka Turcin
Graduate Student at Department of Economics
University of Alaska Fairbanks
Fairbanks, AK 99775.

and

Kelly Giraud
Assistant Professor at Department of Economics
University of Alaska Fairbanks
Fairbanks, AK 99775

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Acknowledgements: The authors would like to thank Ron Dearborn of the Alaska Sea Grant for providing funding and research support for this project. Scott Smiley and Quentin Fong helped with focus group facilitation. John Loomis and Joseph Cooper contributed to survey design. Scott Bates, Stephanie Moreland and Lovro Valcic assisted in survey mailing and data organization. Finally, the authors would like to thank Douglas Larson and John Loomis for their constructive methodological suggestions.

Abstract: This paper examines willingness to pay (WTP) for an endangered species across geographically nested samples using the contingent valuation method (CVM). The samples range from the boroughs that contain critical habitat for the Steller sea lion to the state that contains these boroughs to the entire United States. Depending on the assumptions of the model, WTP varies tremendously from sample to sample. When WTP is restricted to the non-negative region, the WTP does not differ significantly from sample to sample. The estimation results may lead to dramatically different policy implications.

Key words: contingent valuation, geographical nesting, willingness to pay, endangered species, Steller sea lion, Alaska.

Introduction

The contingent valuation method has been used to estimate willingness to pay for a number of rare and endangered wildlife protection policies, including some that involve decisions about species' critical habitat (Loomis and White 1996). It is often useful to use willingness to pay estimates of the United States population for policy decision-making about endangered species. Sometimes, however, it is important to obtain WTP estimates of the region most affected by the potential policy change. While there has been work done on distance from the resource being valued (Sutherland and Walsh 1985, Pate and Loomis 1997), there is very little information in the literature about differences in willingness to pay estimates among geographically nested samples. For example, the counties or boroughs containing the resource being valued, the state where the counties or boroughs are located and the entire United States. This geographical nesting also has a tendency to be nesting of areas, most to least affected by the resource policy change. Studies of effects of geographic distance on WTP estimates in general indicate negative relationship between distance and WTP. In other words, further away respondents are from the resource, the lower is their WTP for the preservation of the resource.

Objective of this paper is to determine whether WTP estimates differ across geographically nested samples (borough, state and national levels) and if they do in which way.

Theoretical Framework

There exist two basic parametric approaches for estimating WTP for public good from CV referendum data. Briefly, one approach develops from utility-maximizing choice model and views referendum data as binary choice data (Hanemann 1984), while the other approach is "based upon the premise that if we could measure valuation exactly, we would use it explicitly in a regression-type model" (Cameron 1988, p. 359). Using either approach yields similar WTP estimates.

Binary Choice Model

Suppose individual n is faced with a choice between two alternatives from a choice set $Cn = \{i, j\}$, where alternative i represents choosing to vote "yes" for tax payment of \$A for public good G and alternative j represents choosing to vote "no" for tax payment of \$A for public good G.

Individual n derives utility U_{in} by choosing alternative i and U_{jn} by choosing alternative j. Following Hanemann (1984), if we define $G = 1$ if i and $G = 0$ if j and using the condition that in consumer equilibrium entire income is spent on gods and/or services, meaning that utility of all other goods can be represented by income I_n, utilities U_{in} and U_{jn} can be formulated as follows:

$$U_{in} = V_{in} + e_{in} = v(1, I_n - A_n, S_n) + e_{in} \qquad (1)$$

$$U_{jn} = V_{jn} + e_{jn} = v(0, I_n, S_n) + e_{jn} \qquad (2)$$

where V_{in} and V_{jn} are assumed nonrandom, systematic components of the U_{in} and U_{jn} respectively, while e_{in} and e_{jn} are assumed random components of the U_{in} and U_{jn} respectively. S_n represents vector of observable attributes of individual n that might affect her/his preferences and A_n represents tax payment of \$A that respondent n can pay for the public good G.

Probability of individual n choosing alternative i is then defined as

$$
\begin{aligned}
P_n(i) &= Pr(U_{in} \geq U_{jn}) \\
&= Pr(V_{in} + e_{in} \geq V_{jn} + e_{jn}) \\
&= Pr\{v(1, I_n - A_n, S_n) + e_{in} \geq v(0, I_n, S_n) + e_{jn}\} \\
&= Pr\{e_{jn} - e_{in} \leq v(1, I_n - A_n, S_n) - v(0, I_n, S_n)\}
\end{aligned} \tag{3}
$$

while probability of individual n choosing alternative j is defined as

$$P_n(j) = 1 - P_n(i) \tag{4}$$

Under the assumption that $e_n = e_{jn} - e_{in}$ is logistically distributed, probability that individual n will choose alternative i can be written as

$$\mathrm{P_n(i)} = \frac{\exp^{\mu V_{in}}}{\exp^{\mu V_{in}} + \exp^{\mu V_{jn}}} = \frac{1}{1 + \exp^{-\mu(V_{in} - V_{jn})}} \tag{5}$$

which is a binary logit model.

Maximum Likelihood Estimates of Binary Logit Model

The log likelihood function of any particular binary choice model can be defined as:

$$L(\beta_1, \beta_2, \ldots, \beta_k) = \sum_{n=1}^{N} \left[y_{in} \log P_n(i) + y_{jn} \log P_n(j) \right] \tag{6}$$

where indicator variable $y_{in} = 1$ if individual n chooses i and $y_{in} = 0$ if individual n chooses j. Note that relationship between and y_{in} and y_{jn} is defined simply as:

$$y_{in} + y_{jn} = 1 \tag{7}$$

If utility is linear in its parameters and if B is defined as a vector of k unknown parameters $\mathrm{B} = [\beta_1, \beta_2, \ldots, \beta_k]$, then

$$\mathrm{P_n(i)} = \frac{1}{1 + \exp^{-BX}} \tag{8}$$

where vector $X = [x_1, x_2, \ldots, x_k]$ consists of k significant explanatory variables in the model. Note that for simplicity μ is set equal to 1.
Binary logit model log likelihood function would then be defined as follows:

$$L(\beta_1, \beta_2, \ldots, \beta_k) = \sum_{n=1}^{N} \left\{ y_{in} \log\left(\frac{1}{1 + \exp^{-BX}} \right) + (1 - y_{in}) \log\left(\frac{\exp^{-BX}}{1 + \exp^{-BX}} \right) \right\} \tag{9}$$

By partially differentiating log likelihood function with respect to each β_k we obtain:

$$\frac{\partial L(\beta)}{\partial \beta_k} = \sum_{n=1}^{N} \{ y_{in} [1 - P_n(i)] - (1 - y_{in}) P_n(i) \} x_{nk}$$

$$= \sum_{n=1}^{N} [y_{in} - P_n(i)] x_{nk} \qquad k = 1, \ldots, K \tag{10}$$

After solving a system of K equations

$$\sum_{n=1}^{N} [y_{in} - P_n(i)] x_{nk} = 0 \qquad k = 1, \ldots, K \tag{11}$$

maximum likelihood estimates of $\beta_1, \ldots, \beta_k$ can be obtained.
If the solution to the first-order conditions exists, it is a unique solution (see Ben-Akiva and Lerman 1985).

Willingness to Pay

If probability of individual n choosing an alternative i is given with

$$\mathrm{P_n(i)} = \frac{1}{1 + \exp^{-(\beta_1 + \beta_2 A_n)}} \tag{12}$$

we can, based on the maximum likelihood estimates obtained by solving (11), estimate mean WTP using the following formula (Hanemann 1989):

$$\textit{unrestricted mean WTP} = \frac{\beta_1}{|\beta_2|} \tag{13}$$

This model implies that mean WTP can assume both positive as well as negative values.

If we wish to rule out negative values of mean WTP, we can truncate the estimate of expected WTP at zero. Given model (12) mean WTP can then be calculated as follows (Hanemann 1989)

$$\textit{restricted mean WTP} = \frac{1}{|\beta_2|} \ln\left(1 + e^{\beta_1}\right) \tag{14}$$

The disadvantage of doing this is that we may overestimate true WTP (Hanemann 1989). It can also be theoretically inconsistent because during the stage of estimating parameters it is assumed that WTP can undertake both negative and positive values. In the stage of calculating mean WTP it is assumed that WTP can undertake only positive values (Haab and McConnell 1997). Nevertheless, this approach is often used in practice as means to solve the "problem" of negative mean WTP.

Case Study

In 1997, after a decline of 80% over 30 years, the western population of the Steller sea lion (*Eumetopias jubatus*) was listed as an endangered species, under the United States Endangered Species Act (Federal Register 62(86): 24345-24355). Reasons for the decline cannot be stated with certainty. However, National Marine Fisheries Service (NMFS)[1], which manages Steller sea lion recovery program, believes that commercial fishing may have contributed to the decline of Steller sea lion. Commercial fisheries may compete for its prey. Even if the commercial fishing industry did not significantly contribute to the decline of the Steller sea lion, restricting harvesting in sensitive areas may help the species recover. Thus, one of the important measures of the program is restricting fish harvesting in the designated critical habitat. Designated critical habitat encompasses areas around Steller sea lion's rookeries, haulout sites and foraging areas in the Aleutian Islands chain, Gulf of Alaska and Bering Sea[2]. In spite of the recovery program, the western population of the Steller sea lion still continues to decline. In November 2000 NMFS issued a set of new Steller sea lion protection measures, which among others consists of more restrictive fish harvesting policy (Federal Register 66(14): 7275-7327).

Contingent Valuation Survey

During summer and fall 2000 a CVM survey was constructed and distributed in order to estimate WTP for the protection program, which would increase restrictions on fish harvesting, and to compare WTP estimates for the program across geographically nested samples. Three samples of 1000 people each where selected from the entire United States, the state of Alaska and the Alaskan Boroughs containing Steller sea lion's critical habitat[3].

The survey was constructed to follow recommendations set forth by the NOAA's blue-ribbon panel (Arrow et al. 1993) as closely as possible. The survey began with an introduction of United States management of the endangered species and its terminology, followed by general opinion questions regarding resource jobs/resource extraction and species protection, which were asked in a likert-scale format[4]. Cummings et al. (1986) refer to this as "researching your preferences". Respondents were also asked about their prior knowledge of Steller sea lions, commercial fishing and coastal Alaskan communities.

A dichotomous choice referendum style question was used to elicit willingness to pay:

> If the Expanded Federal Steller Sea Lion Recovery Program were the only issue on the next ballot and it would cost your household $___ in additional Federal taxes every year for the next ____ year(s), would you vote in favor of it? (By law the funds could *only* be used for the Steller sea lion program.)
>
> YES NO

The bid amounts used in the survey were taken from similar work with endangered species critical habitat protection (Giraud et al. 1999) and were further refined in pre-testing. The payment vehicle was federal taxes. Bid amounts were 1, 3, 5, 10, 25, 50, 75, 100, 200, 350.

In order to investigate temporal elasticity of WTP estimates and to compare it among the samples, each sample was divided and given three treatments. Treatment one was one-year

---

[1] In November 2000 renamed into NOAA Fisheries (National Oceanic and Atmospheric Administration Fisheries), MMPA Bulletin, 2nd/3rd quarter 2000, NMFS, Office of Protected Resources, Silver Spring, MD.

[2] For more see Code of Federal Regulations 50(2): 183.

[3] The services of Survey Sampling, Inc. in Fairfield, CT were employed to obtain a representative sample.

[4] Scale ranged from 1 to 5, where 1 was "strongly disagree", 3 was "neutral" and 5 was "strongly agree".

payment, treatment two was five-year payment and treatment three was fifteen-year payment. This paper does not investigate temporal elasticity and thus, we will present the results pertaining only to the part of each sample containing treatment one.

After the referendum question was a certainty question (a hundred-point line on which respondents were asked to place a mark to express the level of their certainty in the answer to the referendum question) and Yes and No follow-up questions. In the Yes and No follow-up questions, respondents were asked to mark the reasons for voting yes as well as the reasons for voting no. At the very end of the survey, socioeconomic questions were asked.

The Dillman Tailored Design Method (2000) was used for the mailing process.[5] After the third mailing, the overall response rate for this survey treatment was 63.60%. Response rates for *United States*, *Alaska* and *Boroughs samples* are 51.16%, 70.22% and 68.93% respectively.

Results

Logistic Regression

If vector $X = [Bid, ProSpecies, ProJobs, KnowPollock]$, where *Bid* represents price ($A) that respondent can pay in federal taxes for the public good G, *ProSpecies* represents respondent's general opinion on species protection, *ProJobs* represents respondent's general opinion on lost jobs due to species protection and *KnowPollock* represents respondent's knowledge on Pollock fishery in Alaska, then

$$P_n(i) = \frac{1}{1 + \exp^{-(\beta_1 + \beta_2(Bid_n) + \beta_3(ProSpecies_n) + \beta_4(ProJobs_n) + \beta_5(KnowPollock_n))}} \tag{15}$$

Following maximum likelihood estimation of the binary logistic model, these are the coefficient estimates[6] and their z statistics:

Table 1 shows that in all three geographically nested samples, holding everything else constant, coefficients indicate the following: probability of paying $A for the expanded Steller sea lion protection program increases as $A (*Bid*) decreases ($\beta_2 < 0$), the probability increases as respondents' agreement with species protection (*ProSpecies*) increases ($\beta_3 > 0$) and the probability decreases as respondents' agreement with jobs protection (*ProJobs*) increases ($\beta_4 < 0$). Sign of the last coefficient β_5 (*KnowPollock*), is not equal for all three samples. Coefficient $\beta_5 < 0$ for *Boroughs sample*, which indicates that more knowledge respondents from Boroughs have about Pollock fishery in Alaska, less likely they are to pay for the expanded Steller sea lion protection program. On the other hand, $\beta_5 > 0$ for *Alaska sample* and *USA sample*, which indicates that more knowledge respondents from Alaska and from USA have about Pollock fishery in Alaska, more likely they are to pay for the protection program. All signs of the coefficients correspond to the findings of Giraud, Loomis and Johnson (1999), except for the sign of variable *KnowPollock* for *Boroughs sample.*

[5] Announcement letter, 1st survey mailing, reminder postcard, 2nd survey mailing, 3rd survey mailing via Priority Mail.

[6] Coefficients assigned * were significant at $\alpha < 0.01$, coefficients assigned † were significant at $\alpha < 0.05$ other coefficients were significant at $\alpha > 0.05$.

Table 1. Logistic Regression Results

Variable	United States Sample		Alaska Sample		Boroughs Sample	
	Coefficient estimate	z-statistic	Coefficient estimate	z-statistic	Coefficient estimate	z-statistic
Constant	-2.748	-1.369	-2.088	-1.437	0.175	0.131
Bid	-0.009*	-3.555	-0.005†	-2.295	-0.003	-1.440
ProSpecies	1.510*	3.936	1.154*	4.673	0.989*	4.107
ProJobs	-0.841*	-2.632	-0.559*	-2.448	-0.979*	-4.436
KnowPollock	1.961*	2.391	0.089	0.184	-1.169†	-2.008
Number of observations	138		169		178	
pseudo R^2	0.414		0.325		0.334	
Likelihood ratio statistic	78.445		76.143		91.137	

Willingness to Pay for the Expanded Steller Sea Lion Protection Program

Mean WTP estimates, together with their 95% confidence intervals for all three geographically nested samples are shown in Table 2. Unrestricted mean WTP for *United States sample* is \$124.22, for *Alaska sample* it is \$41.20 and for *Boroughs sample* it is \$-208.97. Thus, it could be concluded that willingness to pay increases from *Boroughs* to *Alaska* to *United States sample,* which are also regions most to least affected by the policy change.

This result would indicate positive relationship between WTP estimates and geographically nested samples, which is opposite of relationship between WTP estimates and geographical distance from the resource, established by studies of Sutherland and Walsh (1985) and Pate and Loomis (1997). This may not be a surprising result if we consider differences between the policies and resources investigated in these studies. That can be considered at two levels. First is people's view about the resource and second is their view about the policy. Water or wetlands resources and their quality improvement might be considered more valuable to people living closer to these resources, than might be Steller sea lion and its population recovery for people living in Boroughs. On the other hand, the outcome uncertainty of the Steller sea lion protection policy and lack of knowledge on which policy measures should be implemented, makes this policy less desirable by the people who know more about the policy (i.e. *Boroughs sample*), than by the people who know less about the policy (i.e. *United States sample*). This is supported by the coefficients on *KnowPollock* variable.

The negative sign of Borough's WTP estimate may be debatable. Is it a consequence of statistical fit and functional form or is it an economic harm/negative WTP estimate? Probability of the respondents voting "Yes" is low, even for the smaller bid amounts. The functional form might, thus, be forced to extrapolate into the negative region. On the other hand, people from Boroughs are already experiencing costs of the current policy's fishing restrictions and those costs would only be increased with the expanded protection program. A number of focus group respondents indicated that some fishermen consider Steller sea lions pests because they can interfere with harvests and have been known to jump onto the deck of fishing vessels. Males can weigh up to 2,000 pounds and can be aggressive. Some people from Boroughs might, thus, be against the expanded program even if it were free. They may in effect be including their personal costs when contemplating WTP for the policy.

Restricted mean WTP estimates across geographically nested samples, together with their 95% confidence limits, are shown in Table 2. Restricted mean WTP for *United States sample* is $155.24, for *Alaska sample* restricted mean WTP is $165.01 and for *Boroughs sample* restricted mean WTP is $170.34. Confidence intervals were calculated using a method developed by Park, Loomis and Creel (1991), which is built upon the Krinsky and Robb method (1986). Restricted mean WTP estimates across samples are very close to each other. In order to determine if the difference between the mean WTP estimates are significant, the method of convolutions (Poe, Severance-Lossin and Welsh, 1994) should be applied to this data. When looking at the unrestricted estimates, Borough's WTP decreased drastically, indicating possible mispecification in one of the mean WTP estimates. Furthermore, significant increase in Alaska's WTP suggests that WTP model for Alaska may also be predicting at least some portion of WTP distribution in the negative region, which is supported by the confidence interval of unrestricted mean WTP for this sample.

Table 2. Willingness to Pay across Geographically Nested Samples

	United States Sample	Alaska Sample	Boroughs Sample
Unrestricted mean WTP	$124.22	$41.20	$-208.97
95% confidence interval of mean WTP	[$69.43, $203.64]	[$-140.95, $143.15]	[$-4859.80, $-17.36]
Aggregate Unrestricted WTP	$6.45 Billion	$6.33 Million	$ -5.52 Million
Restricted mean WTP	$155.24	$165.01	$170.34
95% confidence interval of mean WTP	[$110.52, $285.26]	[$97.78, $775.77]	[$79.79, $2367.20]
Aggregate Restricted WTP	$8.06 Billion	$25.36 Million	$4.49 Million

For the policy implications, national level aggregate WTP is of a great importance. Policy makers may also wish to consider values held by state and local residents. Unrestricted and restricted mean WTP estimates for three geographically nested samples are shown in Table 2. One way to estimate aggregate WTP is to simply multiply mean WTP estimates of each sample with number of households in the United States, in Alaska and in the Boroughs. Thus, aggregate unrestricted and restricted WTP shown in Table 2 were calculated based on the 101,562,700 households in the United States, 218,900 households in Alaska and 38,300 households in the Boroughs (Survey Sampling, Inc, 2000). Based on the percentage of non-respondents, we have assumed that 48.86% of the population in the United States has a WTP of zero. For Alaska, we have assumed that 29.78% has a WTP of zero, and for the Boroughs, we have assumed that 31.07% has a WTP of zero.

Aggregate restricted WTP for the United States is 25% higher than the unrestricted WTP. For Alaska, the mean WTP discrepancy is 400%. For the Boroughs, WTP goes from a negative amount ($-5.52 million) to a positive amount ($4.49 million), leading to conflicting policy implications. Caution should be used when interpreting the results presented here for use in policy purposes.

Conclusions

We have shown that restricted and unrestricted mean WTP estimation techniques can lead to dramatically different estimates in terms of policy recommendations. When WTP is not restricted, mean values indicate that the expanded Steller sea lion program results in negative benefits, or additional costs to a segment of the population. The *Boroughs sample* indicates negative mean WTP, while the *Alaska* and *United States samples* show positive WTP when unrestricted. The unrestricted model also suggests positive relationship between WTP estimates and geographically nested samples, which are regions most to least affected by the policy change. Forcing WTP to fall in the non-negative region leads to surprisingly similar estimates of mean WTP. It also results in higher mean WTP estimates for all three samples.

References

Arrow, K., Solow, T., Portney, P., Leamer, E., Radner, R., Schuman, H. 1993. Report of the NOAA Panel on Contingent Valuation, US Department of Commerce, Federal Register 58(10): 4602-4614.

Ben-Akiva, M., Lerman S. R. 1985. "Discrete Choice Analysis, Theory and Application to Travel Demand". The MIT Press. Cambridge, Massachusetts.

Cameron, A. T. 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: 335-379.

Code of Federal Regulations. October 1999. "Fisheries of the Exclusive Economic Zone off Alaska". 50(3) CFR 679: 3437-7815.

Cummings, R. G., Brookshire D. S., and Schulze W. B. 1986. "Valuing Environmental Goods". Rowan & Allenheld, Totowa, New Jersey.

Dillman, D. Mail and Internet Surveys: The Tailored Design Method, 2nd Edition. John Wiley and Sons. 2000.

Giraud, K. L., Loomis, J. B. and Johnson R. L. 1999. "Internal and External Scope in Willingness to Pay Estimates for Threatened and Endangered Wildlife". *Journal of Environmental Management* 56(3): 221-229.

Haab, T. C., McConnell, K. E. 1997. "Referendum Models and Negative Willingness to Pay: Alternative Solutions". *Journal of Environmental Economics and Management* 32(2): 251-270.

Haab, T. C., McConnell, K. E. 1998. "Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay". *Land Economics* 74(2): 216-229.

Hanemann, M. W. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses". *American Journal of Agricultural Economics* 66(3): 332-341.

Krinsky, I., Robb, A. (1986), Approximating the Statistical Properties of Elasticities. Review of Economics and Statistics 68(Nov): 715-719.

Loomis, J. B., White, D.1996. "Economic Benefits of Rare and Endangered Species: Summary and Meta Analysis". *Ecological Economics* 18: 197-206.

Federal Register. January 2001. "Steller Sea Lion Protection Measures for the Groundfish Fisheries Off Alaska". Federal Register 66(14): 7275-7327.

Federal Register. May 1997. "Threatened fish and wildlife; change in listing status of Steller sea Lions under the Endangered Species Act". Federal Register 62(86): 24345-24355.
Pate, J., Loomis, J. 1997. "The effect of distance on willingness to pay values: a case study of wetlands and salmon in California". *Ecological Economics* 20(3): 199-207.
Park, T., Loomis, J., and Creel, M. (1991), Confidence Intervals for Evaluating Benefits from Dichotomous Choice Contingent Valuation Studies. *Land Economics* 67(1):64-73.
Poe, G., Severance-Lossin, E., and Welsh M. (1994), Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach. *American Journal of Agricultural Economics* 76(4): 904-915.
Survey Sampling, Inc. (2000). Sample Specification Report. Fairfield, CT.
Sutherland, R. J., Walsh, R. G. 1985. "Effect of Distance on the Preservation Value of Water Quality". *Land Economics* 61(3): 281-291.

The Infeasibility of Wetland Restoration to Reduce Flood Damage in the Red River Valley: Would the Inclusion of Non-Market Wetland Values Make a Difference?

Dr. Steven D. Shultz[1]

And

Dr. Jay A. Leitch[2]

North Dakota State University

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Introduction

A previous study found that it was not economically feasible to restore wetlands in order to reduce flood damage in at least one watershed of the Red River Valley of the North (Shultz, 1999). Following that study critics have contended that the inclusion of non-market wetland values in the aforementioned cost-benefit would result in wetland restoration being considered feasible. This present paper evaluates such a claim by reviewing previous non-market wetland valuation studies in the region and by evaluating whether the inclusion of these values would influence feasibility measures.

A Review of the Original Study of Wetland Restoration for Flood Control

A multi-disciplinary study funded by the International Joint Commission (IJC) after the extensive flooding in the Red River Valley (RRV) in 1997 focused on if it was economically feasible to restore previously drained wetlands in North Dakota's Maple River Watershed (Shultz, and Leitch, 2000). The Maple River Watershed is a typical RRV, agricultural-based watershed with many acres of both existing and drained wetlands, and frequent springtime flooding events both within and downstream of the watershed.

An evaluation of the 2,700 acres of previously drained wetlands in the middle and upper reaches of the Maple River watershed found 700 acres of small (less than 1 acre) wetlands, 1,100 acres of medium (1 to 5 acre) wetlands, and 900 acres of large (greater than 5 acre) wetlands. Accounting for these size distributions, hypothetical construction costs were calculated to be $520,000 for simple restoration, $1,040,000 for restoration with outlet controls, and $3,650,000 for complete restoration. Average land rental values based on the location of previously drained wetlands were estimated to be $35 an acre per year and $26.5 per acre for low cost land. Therefore, the annual land rental cost for restoring 2,700 acres of wetlands in the watershed is $108,000 and $94,500 for low cost land.

The period of 1989-1998 was considerably wetter than the historical 50-year average. During this 10-year time period this time period there were 4 high frequency (2-year, 50 percent probability) floods, 1 medium frequency (10-year, 10 percent probability) flood, 2 low frequency (25-year, 4 percent probability) floods and a 1 very low-frequency (50-year, 2 percent probability) flood. All of these flood events combined resulted in $29.3 million in damage within the Maple River Watershed of which about half was agricultural damage. The 1997 flood event was contributed to the $2 billion in damage downstream in the City of Grand Forks of which approximately 7 percent ($140 million) can be attributed to flows from the Maple River watershed. Therefore, total flood damage associated with the Maple River Watershed from 1989 to 1998 is $169.3 million measured in 1998 dollars

Accounting for the probability of observed (1989-1998) floods occurring in any particular year and expected reductions in peak flood stages and hence damages, the average annual avoided flood damage associated the hypothetical restoration of 2,700 wetland acres is $9,950 with 1 foot of storage bounce or $14,500 with 2 feet of bounce. The present value of these benefits over a 20-year period using a 5 percent discount rate is $124,000 with 1 foot of bounce and $181,000 with 2 feet of bounce.

All of the wetland restoration alternatives evaluated have very large negative net present values ranging from - $1.5 million to - $ 4.6 million with corresponding benefit-cost ratios from 0.04 to 0.11 which unequivocally indicates economic infeasibility for each of the wetland restoration options (Table 1). As expected, the more costly outlet and complete restoration options are the least feasible alternatives despite their relatively higher avoided flood damage benefits. Even the most optimistic (and unlikely) scenario of simple restoration on low cost land and with 2 feet of storage bounce is not feasible with costs exceeding benefits nine fold.

Previous Non-Market Valuation Wetland Studies in the Region

Wetlands are seen as providing many valuable services to society such as improved water quality, groundwater recharge, the support of fish, wildlife and plant habitat, recreational and amenity services, and flood control (Mitsch and Gosselink, 1993). However, because many of the services that wetlands provide are not actively traded in the market, determining their economic value has been difficult (Whitehead, 1993, and Leitch and Ludwig, 1995).

Hedonic based valuations of the amenities of wetlands in urban areas based on the estimation of statistical relationships between site-specific wetland characteristics and local housing values have been quite successful (Doss and Taff, 1996, and Mahan, Polasky and Adams, 2000). Both recreational (use) and existence (non-use) values of existing wetlands have been estimated through combinations of the contingent valuation and travel cost methods, but transfer of these site-specific estimates to other areas may be limited (Bergstrom et al., 1990 and Whitehead 1993)

In and around the RRV, only two site-specific wetland valuation studies have been conducted despite the fact that the Prairie Pothole Region is one of the largest concentration of wetland resources in the 48 continuous states.

The first study estimated the societal values of four typical prairie pothole wetlands (including one in the Maple River Watershed), and one wetland complex, all in North Dakota (Hovde and Leitch, 1994 and Leitch and Hovde, 1996). The average annual value for flood control of $2.5 per acre was estimated by calculating the equivalent storage costs of nearby flood control projects (usually retention basins). The average annual value for reducing sedimentation of $0.05 per acre was estimated by calculating avoided drainage ditch excavation resulting from sediments being trapped by wetlands. Recreation and wildlife related values of $4.12 per acre were made by calculating annual expenditures associated with the use of wetlands in addition to a 40 percent consumers' surplus premium. No estimates of the aesthetic or existence (non-use) values of the wetlands were made because of the abundance of wetlands in comparison to people within and nearby the study areas. Similarly, no estimates were made of the groundwater recharge values because groundwater was not utilized as a water source in the areas studied.

The second study by Roberts and Leitch (1997) found that Mud Lake reservoir and wetland complex in the southern end of the RRV contributed 57 percent of avoided downstream historic flood damage which was converted to a value of $440 per acre annually. This same study also used the contingent valuation method to determine that the combined habitat, recreation, and aesthetic values for the wetland complex were $21 per acre per year. Unfortunately, as the authors mention, the characteristics of the Mud Lake wetland complex are more similar to a managed reservoir than the smaller and more ubiquitous prairie pothole wetlands in the region, which limits the transfer these valuation results to other locations.

The Inclusion of Non-Market Wetland Values in the Feasibility Analysis.

The value of non-flood related wetland benefits required for alternative wetland restoration options to break-even range from $28 to $86 per acre per year (Table 1). The higher end values of $83 and $86 associated with complete restoration options are of the most interest because these restoration alternatives are specifically designed to provide such wetland values. However, these required break-even values greatly exceed previously reported average annual per acre wetland benefits of $4.2 for sedimentation control and recreation among five nearby wetlands (Leitch and Hovde, 1996), as well as the combined recreation and existence value of $21 for a RRV wetland complex (Roberts and Leitch, 1997), and the $5 to $9 range of recreation-based wetland values in Louisiana (Farber 1987 and Bergstrom et. al., 1990).

There are several reasons why non-market wetland values are relatively low in the in both Red River Valley and North Dakota as a whole. First there are large numbers of existing wetlands especially in relation the low population of North Dakota. Specifically there are approximately 650,000 people in North Dakota and over 2.7 million wetland acres (4 wetland acres for every person in the State). Second very few North Dakotan's live or recreate on these wetlands, with the exception of waterfowl hunters. Finally, waterfowl species in the region that utilize wetland habitat for breeding are doing extremely well due to the extended wet-period in the region. Also, the critical waterfowl habitat is most often associated with shallow temporary wetlands rather than the deeper wetlands, which are restored for flood control purposes.

Table 1. Feasibility Indicators of Wetland Restoration Options and Required Non-Market Wetland Values in Order To Break Even

Wetland Restoration Option	**Net Present Value**	**Benefit Cost Ratios**	**Non-Market Wetland Values Required to 'Break- Even' (Acre/Year)**
Simple Restoration (1 ft bounce, 2,700 AF storage)	- $1.7 million	0.07	$32
Simple Restoration (Low Cost Land) (1 ft bounce & 2,700 AF storage)	- $1.5 million	0.07	$29
Outlet Restoration (2 ft bounce & 5,400 AF Storage)	- $2.1 million	0.08	$40
Outlet Restoration (Low Cost Land) (2 ft bounce & 5,400 AF Storage)	- $1.9 million	0.08	$37
Complete Restoration (2 ft bounce & 5,400 AF Storage)	- $4.6 million	0.04	$86
Complete Restoration (Low Cost Land) (2 ft bounce & 5,400 AF Storage)	- $4.4 million	0.04	$83
Simple Restoration, Low Cost Land with Maximum Storage Bounce (2 ft of bounce, & 5,400 AF Storage)	- $1.5 million	0.11	$28

In conclusion, rather than undertaking large-scale wetland restoration projects and/or conducting non-market valuation studies of wetlands in North Dakota, scarce resources should be directed to identifying strategies to cost-effectively preserve existing wetlands. Particular emphasis should be place on making optimal conservation easement purchases based on the location of low cost land, high priority wildlife habitat, and sub-watershed level flood control needs.

References

Bergstrom, J.C., J.R Stoll, J.P Titre and V.L. Wright. Economic Value for Wetlands-Based Recreation. *Ecological Economics*, 2:129-147. 1990

Doss, C.R. and S.J. Taff. The Influence of Wetland Type and Proximity on Residential Property Values. *Journal of Agricultural and Resource Economics.* 21(1):120-29. 1996.

Farber, S. The Economic Value of Wetland Systems. *Journal of Environmental Management*, 24:41051 1987.

Hovde, B. and J.A. Leitch. Valuing Prairie Potholes: Five Case Studies. *Agricultural Economics Report* # 310. North Dakota State University. 1994.

Leitch, J. A., and B. Hovde. "Empirical valuation of prairie potholes: Five case studies." *Great Plains Research* 6(1996):25-39

Mahan, B.L. S. Polasky, and R.M. Adams. Valuing Urban Wetlands: A Property Price Approach. *Land Economics.* 76(1):100-113. 2000

Mitsch, W. J. & J.G. Gosselink. Wetlands, 2nd ed. New York. Van Nostrand Reinhold. 1993.

Roberts, L. A., and J. Leitch. "Economic valuation of some wetland outputs of Mud Lake Minnesota-South Dakota." Departmental Report #381, Department of Agricultural Economics, North Dakota State University, Fargo. 1997.

Shultz, 1999. The Feasibility of Wetland Restoration to Reduce Flooding in the Red River Valley: A Case Study of the Maple (ND) and Wild Rice (MN) Watersheds. Agricultural Economics Report # 23, North Dakota State University.

Whitehead, J.C. Economic Valuation of Wetland Resources: A Review of the Value Estimates. *Journal of Environmental Systems* 22(2):151-161. 1993

Testing the Validity of Benefit Transfers: A Site Correspondence Model

Randall S. Rosenberger
Assistant Professor
Regional Research Institute and Division of Resource Management
West Virginia University

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Acknowledgements: Data used in this analysis was graciously provided by John Loomis, Jeffrey Englin, and Armando Gonzalez-Caban. The manuscript benefited from comments provided by the friends and members of the Western Regional Research Group W-133, Benefits and Costs Transfer in Natural Resource Planning, at the annual meetings in Miami, FL, 2001. Any shortcomings of the analysis are the sole responsibility of the author. Send all correspondence to: Randall S. Rosenberger, P.O. Box 6108, 2032 Agricultural Sciences Building, West Virginia University, Morgantown, WV 26506-6108, Phone: (304) 293-4832 x4468, FAX: (304) 293-3752, rrosenbe@wvu.edu

Abstract: Several factors can affect the validity and reliability of benefit transfers. This paper uses regression analysis to investigate the systematic effects of differences in sample and site characteristics on the magnitude of error associated with an experimental benefit transfer. Validity measures are derived through various specifications of multi-site and single-site travel cost demand models for hiking on a variety of trails in Colorado. The results show that some characteristics account for a large portion of error in the benefit transfer application. Meta-analysis is used to develop a calibrated benefit transfer function that results in more accurate and reliable transfer measures. (JEL Q20)

Key Words: Benefit Transfer, Hiking, Meta-Analysis, Site Correspondence, Validity

Introduction

Benefit transfer is the adaptation and use of existing information or data to new contexts. Benefit transfer has become a practical way to inform decisions when primary data collection is not feasible due to budget and time constraints, or when expected payoffs are small. For our present purpose, benefit transfer is the adaptation and use of economic information derived from primary research on non-market values under certain resource and policy conditions. This information is then used to inform management and policy decisions for similar issues and conditions as the primary research context. Traditionally, the context of primary research is referred to as the study site, and the benefit transfer context is referred to as the policy site.

Benefit transfer may be used to inform decisions at various stages. It may be used to inform the framing of issues, to identify the extent of a market, as a pre-screening of natural resource damage assessments, to justify the collection of primary data, or for other facets of the decision process. Benefit transfer is not limited to predicting a value measure for assessing costs and benefits, although this is a primary use.[1] Instead, benefit transfer adapts our accumulated body of knowledge based on evidence from primary research and past experiences to identify bounds on the scope of potential outcomes. For this reason, benefit transfer may be as important in policy evaluations as conducting primary research.

Benefit transfer, however, is not a panacea for all valuation needs. Benefit transfer provides content- and context-relevant value measures at best; whereas primary research provides content- and context-specific measures of value.[2] The outcomes of primary research, unfortunately, may be treated as a public good with an incentive for agencies to free-ride on past knowledge.[3] Information obtained through primary research is part of the process of accumulating knowledge about the value of non-market resources, and thus each new effort in primary research adds to our depreciating stock of knowledge.[4] The quality of benefit transfers relies on the quality of this body of knowledge. Thus, the functional value of benefit transfers is bound by primary research. The polar extreme, or worst-case scenario, is to ignore non-market values, implying these resources have no value in an economic evaluative framework.

Several benefit transfer methods have been used and tested. Two categories of benefit transfers can be identified in an inventory of the different methods (Rosenberger and Loomis, forthcoming). The first category is value transfers. These methods include the transfer of a single measure, a range of point measures, measures of central tendency, or the application of agency approved measures. The second category is function transfers. These methods include the transfer of demand, benefit, or willingness to pay functions and meta-analysis functions.

Validity measures reported in past tests of the accuracy of benefit transfers illustrate the potential variability of these measures (Table 1). Most empirical tests of benefit transfers culminate with these validity measures. However, little research has been conducted on the relationship between these measures and the factors that can affect them.[5] These factors can include the quality and robustness of the data, the methods used in modeling and interpreting the data, and the correspondence between the study site and the policy site. Several protocol for conducting benefit transfers have been suggested in the literature as an attempt to minimize the effect of these factors on the error associated with benefit transfers (Rosenberger and Loomis 2001).

This paper presents the results of a project that goes beyond the traditional tests of validity and reliability by relating the validity measures to site correspondence factors and developing a meta-analysis benefit transfer function. The structure of the paper is as follows. The next section describes the models and methodology used, followed by a description of the data. The multi-site and single-site travel cost models are presented next, including model specification, consumer surplus measures, and benefit transfer validity measures. These validity measures are then related to measures of the correspondence between the sites. A meta-analysis benefit transfer function is developed and applied, illustrating the gains in accuracy over the traditional benefit transfer methods used. The paper ends with a discussion regarding some of the results and issues encountered in this research.

Site Correspondence Model Development

This analysis builds on a simple point estimate value transfer approach. Primary research estimates a study site benefit measure for site i, V_{Si}. This measure is then transferred to a different, but similar site j. Site j is the policy site. When the study site measure, V_{Si}, is transferred to the policy site, it becomes a transfer value, V_{Tj}. However, we do not know the actual value (V_{Pj}) for policy site j. V_{Ti} approximates V_{Pi}, or

$$V_{Ti} = V_{Pj} + \delta_{ij}, \quad [1]$$

when $i \neq j$. δ_{ij} is the error associated with the transfer of a benefit measure from site i to site j. Measures of V_{Ti} and V_{Pj} are derived from travel cost models of demand for recreation trips.

Travel Cost Models

Each individual is assumed to have preferences for the consumption of market goods and recreation trips, as represented in a utility function, U(.). It also is assumed that each individual attempts to maximize his/her utility ([2]) by allocating scarce (constrained; i.e., [3] and [4]) resources to a variety of consumption and recreation choices (Fletcher, Adamowicz, and Graham-Tomasi 1990):

$$\max U_i(\mathbf{x}_i, \mathbf{v}_i, \mathbf{L}_i; \mathbf{q}_i) \quad [2]$$

subject to:

$$\mathbf{p}\mathbf{x}_i + c_i\mathbf{v}_i \leq Y_i$$
(money budget constraint), and [3]

$$T^w_i + T^L_i + T(\mathbf{v}_i) = T \text{ (time budget constraint)}, \quad [4]$$

where x, v, and q are vectors of market goods, recreation visits, and recreation site qualities for individual i. L_i is leisure time spent in recreation activities other than activity v. The money budget constraint restricts money expenditures for market goods **x** at prices p and for recreation trips **v** at travel costs c. T is the time budget with T^w being time spent at work, T^L being time spent in leisure activities other than v, and $T(v)$ being the duration and frequency of time spent in recreation activity v.

Maximizing [2] subject to the constraints ([3] and [4]) results in vectors of demand functions for market goods **x** and recreation activities **v**. Assuming separability in choices across all cost, price, income and quality parameters, the following demand function is identifiable for recreation activity v:

$v_i\ (c_i, q_i, Y_i, T(v_i))$ [5]

The demand for recreation activity v is a function of travel costs (c) including entrance fees, the perceived quality of the recreation site (q), income (Y), and time spent in recreation activity v ($T(v)$).

Validity Measures

Empirical tests of the convergent validity of benefit transfers estimate V_{Pj} through primary research, or the case where $V_{Si} = V_{Pj}$ when i = j. Other estimates derived from primary research are used as transfer values to the policy site, $V_{Si} = V_{Ti}$. Convergent validity measures are calculated as percent differences between V_{Ti} and V_{Pj} as

$$\%\Delta V_{ij} = [(V_{Ti} - V_{Pj})/V_{Pj}] * 100 \quad [6]$$

when $i \neq j$. Given equations [1] and [6], the convergent validity measures become $\delta_{ij}/V_{Pj} * 100$.

Site Correspondence Model

The correspondence between the study site and policy site is arguably a contributing factor in the validity or accuracy of benefit transfers (Berglund, Magnussen, and Navrud 1995; Boyle and Bergstrom 1992; Brouwer 2000; Desvousges, Naughton, and Parsons 1992). We investigate the extent to which differences in the physical attributes of the sites systematically contribute to the magnitude of the validity measures. Evidence from tests using multi-site models suggests intra-state transfers outperform inter-state transfers where the variability in attributes is more likely (Downing and Ozuna 1996; Kirchhoff, Colby and LaFrance 1997; Loomis 1992; Loomis et al. 1995; VandenBerg, Poe and Powell 2001). Other factors that may contribute to the validity of benefit transfers include the quality of the data for the study site and the quality of methodology used. By holding data and methodological factors constant, this analysis focuses on the effect of site difference on the magnitude of validity measures.

A site correspondence model can be defined as

$$\%\Delta V_{ij} = f(\%\Delta \text{Market}_{ij}, \%\Delta \text{Site Characteristics}_{ij}). \quad [7]$$

That is, the percent difference in the value transfer ($\%\Delta V_{ij}$) is a function of the percent difference in the markets (($\%\Delta \text{Market}_{ij}$) and the percent difference in the characteristics of the sites ($\%\Delta \text{Site Characteristics}_{ij}$), where i is the study site and j is the policy site. Market characteristics can be measured in terms of the demographic profiles of the affected populations for the sites and the site characteristics can be measured as physical differences between the sites.

Meta-Analysis Model

If different factors are related to the degree of accuracy in benefit transfers, then it may be possible to develop a transfer function that calibrates measures based on these identifiable differences. This calibrated transfer function, in theory, should provide more valid and reliable transfer measures. Meta-regression analysis may be a potential method for developing this function. Meta-analysis is the statistical summarizing of relationships between benefit measures and quantifiable characteristics of studies. Meta-analysis has been traditionally concerned with understanding the influence of methodological and study-specific factors on research outcomes and providing summaries and syntheses of past research (Rosenberger and Loomis 2000b). A more recent use of meta-analysis is the systematic utilization of existing value estimates from the literature for the purpose of benefit transfer (Desvousges, Johnson, and Banzhaf 1998; Kirchhoff 1998; Sturtevant, Johnson, and Desvousges 1998; Rosenberger and Loomis 2001). One potential advantage of using meta-analysis functions for benefit transfer is the increased sensitivity of transfer estimates to characteristics of the policy site.

A meta-analysis benefit transfer function can be defined as:

$$V_{Si} = f(\text{Market}_i, \text{Site Characteristics}_i) \quad [8]$$

That is, the benefit measure derived for site i is a function of its market characteristics and its physical attributes. We can test whether the meta-analysis benefit transfer function provides more accurate benefit estimates for a policy site than traditional benefit transfers. The hypothesis test is:

$$H_0: \ \%\Delta V_{METAij} = \%\Delta V_{TRADij} \quad [9]$$

$$H_1: \ \%\Delta V_{METAij} < \%\Delta V_{TRADij}. \quad [10]$$

$\%\Delta V_{METAij}$ is calculated in the same fashion as $\%\Delta V_{ij}$ ([6]). A one-tailed paired t-test may be used to test the null hypothesis that the meta-analysis transfer approach and the traditional transfer approach applied herein are equally accurate versus the alternative hypothesis that the meta-analysis transfer is more accurate than the traditional approach.

Data

Survey Design

The data used in this analysis was collected in 1998 for the purpose of investigating the effects of forest fires on the value of hiking and mountain biking (Loomis, Gonzalez-Caban, and Englin 2000). Trails were selected in a stratified random sampling of past fire occurrences by age of fire and acres burned. Recreation users of the trails were sampled during July and August. Over a 35 day period, 10 trails were sampled on a weekday and a weekend day by intercepting recreation users as they returned to a trailhead parking area. They were provided with a statement regarding the purpose of the survey and a mail-back questionnaire. The questionnaire elicited information from the users about their primary activity on the trail, travel cost information (travel time, distance from home to the trail, travel costs, number of trips this year and last year to the current trail), and sociodemographics, among other questions. Table 2 provides a description of the variables used in this analysis. A total of 527 surveys were distributed with 354 being returned, for a response rate of 67 percent. The data used in this analysis is restricted to hikers and trails predominantly used by hikers. Therefore, the current sample consists of 127 respondents across six individual or combined trails in three National Forests. Some of the trails were combined based on similarity of characteristics and proximity to each other in order to improve the degrees of freedom in the analysis.

Three National Forests (NF) in Colorado (the Arapaho-Roosevelt NF, the Gunnison-Uncompaghre NF, and the Pike-San Isabel NF) were selected, providing a range of fire and trail characteristics (Table 3). Two of the National Forests are along the front range of the Rocky Mountains with the other National Forest being interior to the mountains. The sample of 127 recreational hikers also provides a range of demographic characteristics (Table 4).

Travel Cost Models

Several travel cost models are estimated from different levels of the data. The models estimated include (1) state-wide, multi-site travel cost models based on different travel cost (TC) specifications; (2) forest-specific, multi-site travel cost models based on different TC specifications; (3) N-1 travel cost models[6]; and trail-specific travel cost models that serve as the underlying demand for that trail. All models are estimated as negative binomial random effects models. The negative binomial model is statistically appropriate given the dependent variable is non-negative integer, count data (Creel and Loomis 1990; Englin and Shonkwiler 1995) and is overdispersed (mean and variance are not equal) (Englin and Cameron 1996). The random effects specification is used to account for the panel nature of the data, i.e., multiple responses from each individual (trips this year and trips last year) (Rosenberger and Loomis 1999, 2000a).

The basic form of the travel costs models [5] is given by:

$$\ln Trips_{ijk} = \alpha + \beta TC_{ijk} + \gamma Demographics + \varphi Site\ Characteristics_k + \mu_{ij} + \varepsilon. \qquad [11]$$

The dependent variable is number of hiking trips this year (i) and last year (j) to site k and is naturally logged in the negative binomial specification. Trips is a function of travel costs to site k this year (i) and last year (j), demographics of the respondent, and characteristics of site k. α is the intercept term and β, γ, and φ are parameters to be estimated. μ_{ij} is the random panel effect and ε is the common error component.

State-Level and Forest-Level Multi-Site Models

Table 5 provides the estimated travel cost models for state-level and forest-level specifications. Model A pools the data and specifies a common travel cost (*TC*) variable. This model results in a ballpark demand model for hiking trips in the state. Model B pools the data and specifies forest-specific travel cost shift variables ($TC*F_i$). Model C pools the data and specifies trail-specific travel cost shift variables ($TC*T_i$). Model D pools the data for the Arapaho-Roosevelt National Forest and specifies a common travel cost variable (*TC*) for the forest. Model E pools the Arapaho-Roosevelt data and specifies trail-specific travel cost shift variables ($TC*T_i$). Model F pools the Gunnison-Uncompaghre data and specifies a common travel cost variable (*TC*) for the forest. Model G pools the Gunnison-Uncompaghre data and specifies trail-specific travel cost shift variables ($TC*T_i$) for the forest. Several of the demographic and site characteristic variables have estimated covariates of different signs. These variables affect number of trips in opposite directions. Adapting some of these models to other recreation sites may result in value estimates that are adjusted in the wrong direction. That is, adjusting a model for a positive covariate effect at the study site when the effect is inverse at the policy site.

N-1 Multi-Site Models

Table 6 provides the estimated travel cost models for the N-1 multi-site modeling of the data. Loomis (1992) and VandenBerg, Poe and Powell (2001) use an N-1 data-splitting modeling approach to derive transferable benefit measures. This approach provides general benefit measures by pooling all of the data except the n^{th} site, and then uses the $n\text{-}1^{th}$ benefit measure as the transfer measure for the n^{th} site. Model H pools all of the data except for Trail 1, Model I pools all of the data except for Trail 2, and so forth up to Model M that omits data for Trail 6. These models most closely resemble Model A, the state-level multi-site model with a common *TC* variable. The covariate estimates on *TC*, *GENDER*, and *EDU* are consistent with Model A. However, the covariate estimates on *LONG*, *FIRE*, and *WATER* are the opposite of the covariate estimates on these variables for Model A. This implies the effect of these variables is sensitive to subsets of the data.

Trail-Specific Models

Table 7 provides the estimated travel cost models for each trail. These travel cost models provide the demand function specific to each trail. They also will serve as the baseline models when estimating V_{Pj} (actual benefits) for each trail. Note that none of the site characteristic variables is present in these models. This is because the measures of site characteristics are invariant intra-trail. Variation in site characteristics is only available for inter-trail models.

Benefit Measures

The benefit estimates derived from the travel cost models are measures of Marshallian consumer surplus. Consumer surplus is calculated by integrating the demand function (equation 5) over the relevant price or travel cost range, yielding consumer surplus per trip. The negative binomial random effects model is equivalent to a semi-log demand function. Therefore, consumer surplus can be simply calculated as $(-1/\beta_i)$, or -1 divided by the coefficient on travel cost (Adamowicz, Fletcher, and Graham-Tomasi 1994; Creel and Loomis 1990). In those models with forest- or trail-specific travel cost shift variables, the formula is $[-1/(\beta_1+\beta_2)$. For example, consumer surplus for Trail 1 in Model C is calculated as $[-1/(-0.0153+0.0074)]$.

Table 8 provides the benefit measures for each of the travel cost models. The baseline measures are calculated from the trail-specific travel cost models (Table 7) and provide estimates of V_{Pj} for equation [1]. The transfer measures of equation [1] (V_{Ti}) are calculated from Model A through Model M (Tables 5 and 6). Table 8 shows that there is a wide range of benefit measures from a low of $12.12 per trip for Trail 1 to a high of $248.85 per trip for Trail 4. Table 8 also shows that there is an increase in the variability of hiking values as we progress from Model A (which provides a ballpark estimate of the value of hiking in Colorado), to more specific measures for a National Forest (Models B, D and F), to the value of hiking for a specific trail (Models C, E, G, and H through K).

Validity Measures

Table 9 provides the validity measures as percent differences ($\%\Delta V_{ij}$) ([6]) between the transfer value (V_{Ti}) and the actual, or baseline value (V_{Pj}). These measures provide an indication of the relative accuracy of the benefit transfer process when V_{Pj} is known.[7] The validity measures range from a low of about 4 percent underestimating the value of Trail 6 using Model G to a high of over 900 percent overestimating the value of Trail 1 using Model C. The average percent difference measures for each model ranged from an average of about 43 percent when using the forest-specific travel cost models with trail-specific travel cost shift variables (Models E and G) to over 200 percent when using the n-1 modeling strategy (Models H through M). These measures are consistent with other empirical measures from the literature (Table 1).

Site Correspondence Model

This section defines the site correspondence model ([7]). The validity measures presented in Table 9 form the dependent variable in the site correspondence model ([7]). The explanatory variables are calculated in a similar fashion by applying equation [6] where V_{Ti} becomes demographic and site characteristic measures for site i, and V_{Pj} becomes the corresponding measure of the characteristic for policy site j. The specific form of the site correspondence model is:

$$\%\Delta V_{ijm} = \alpha + \beta_1\%\Delta GENDER_{ij} + \beta_2\%\Delta AGE_{ij} + \beta_3\%\Delta ELEV_{ij} + \beta_4\%\Delta GAIN_{ij} + \beta_5\%\Delta LONG_{ij} + \beta_6\%\Delta FIRE_{ij} + \beta_7\%\Delta WATER_{ij} + \beta_8\%\Delta PP_{ij} + \beta_9\%\Delta LP_{ij} + \beta_{10}\%\Delta ASP_{ij} + \mu_m + \varepsilon. \quad [12]$$

This model investigates the magnitude of the effect of differences in market and site characteristics between study site i and policy site j on the error associated with the transfer of a benefit measure from site i to site j using modeling strategy m (state-level, forest-level, or N-1 modeling). In this model, μ_m is the panel-specific error component and ε is the common error component.

The dependent variable ($\%\Delta V_{ijm}$) is of the panel data type; multiple observations are from the same source (modeling strategy). Identifying the strata or panels is an important component when dealing with panel data (Rosenberger and Loomis 2000a). In this case, the modeling strategy m is a potential source of panel effects. Three unbalanced panel strata are defined: (1) 18 validity measures are derived from applying the State-level Models A through C (Table 5), (2) ten validity measures are derived from applying the Forest-level Models D through G (Table 5), and (3) six validity measures are derived from applying the N-1 Models H through M (Table 6). A random effects generalized least squares regression technique is used because some of the regressors are invariant within a panel. A fixed effect estimator requires all regressors to have an intra-panel variance.

Table 10 provides the results of the fitted site correspondence model. We have no expectations regarding sign and significance of the explanatory variables. The model has an adjusted-R^2 of 0.79. Interpretation of the estimated covariates is relatively straightforward. First, a significant variable in the regression indicates the variable has an effect on the accuracy of benefit transfers. Second, because the variables are unitless measures of percent difference between the study site and the policy site ([6]), a positive (negative) sign indicates that this variable leads to an overestimation (underestimation) of the policy site's baseline value. Third, the larger the coefficient on each variable, the greater the effect on the accuracy of the benefit transfers.

Table 10 shows that the demographic variables ($\%\Delta GENDER$, $\%\Delta AGE$) have a significant and positive effect on the accuracy of the benefit transfers, with differences in gender composition of the samples having over twice the effect as differences in the age composition of the samples. Differences in the lengths of the trails ($\%\Delta LONG$) result in an overestimation of transferred values. In descending order of the magnitude of their effect, differences in elevation ($\%\Delta ELEV$), presence of water ($\%\Delta WATER$), and gain in elevation of the trail ($\%\Delta GAIN$) have a negative effect on the accuracy of the benefit transfers. Tree cover type of the recreation sites, in particular the presence of lodgepole pine ($\%\Delta LP$) and aspen forests ($\%\Delta ASP$), have somewhat significant effects on the accuracy of the benefit transfers by underestimating and overestimating the target policy value, respectively.

Meta-Analysis Transfer Function

The final step in this analysis is to define a meta-analysis benefit transfer function for hiking values in Colorado's National Forests ([8]). The justification for developing a meta-analysis transfer function is because of the wide-variability in the accuracy of the traditional value transfers (Table 9) and evidence presented that this variability is due to differences between the study site and the policy site (Table 10). The meta-analysis function acts as a calibration of the benefit measures to characteristics of the sites. This procedure is similar to Feather and Hellerstein's (1997) calibration approach. However, instead of calibrating a transfer function to a fixed, arbitrarily defined target, the transfer function is developed using regression analysis and moving targets. The targets are moving in this approach because the baseline measures (Table 8) derived from the trail-specific models (Table 7) are themselves estimates of an unknown value. Therefore, the dependent variable in the regression analysis is composed of all benefit measures reported in Table 8.

The data for the meta-analysis function also is of the panel data type. The strata are the same as the site correspondence model above with the exception that there are now four strata: (1) 18 observations derived from the State-level Models A through C (Tables 5 and 8); (2) ten observations derived from the Forest-level Models D through G (Tables 5 and 8); (3) six observations derived from the N-1 Models H through M (Tables 6 and 8); and (4) six observations derived from the Trail-specific models (Tables 7 and 8). A random effects generalized least squares regression technique is used to fit the data. A fixed effects specification is inappropriate because some of the panels are invariant in some of the regressors.

The specific form of the empirical model is:

$$CS_{im} = \alpha + \beta_i TRAIL_i + \beta_6 ELEV_m + \beta_7 GAIN_m + \beta_8 LONG_m + \beta_9 FIRE_m + \beta_{10} PP_m + \mu_m + \varepsilon. \quad [13]$$

That is, consumer surplus (CS_{im}) for the i^{th} trail using the m^{th} travel cost model is a function of $TRAIL_i$ (a dummy variable identifying the trail where i is Trail 1 through Trail 6 (Table 3) with Trail 2 being the omitted variable), and site characteristic measures for elevation ($ELEV_m$), gain in elevation ($GAIN_m$) and length ($LONG_m$) of the trail, age of past forest fires ($FIRE_m$), and ponderosa pine forest type (PP_m) for the m^{th} model. μ_m is the panel-specific error component and ε is the common error component. Several of the variables could not be included in the model because they were either correlated with μ_m (e.g., the demographic variables) or correlated with the trail-specific dummy variables (e.g., WATER is a characteristic of Trail 3).

Table 11 provides the results of the estimated meta-analysis benefit transfer function. The adjusted-R^2 of the model is 0.72. Although only two of the variables (*Trail 3* and *LONG*) are significant at the 0.10 level or better, the majority of the other variables are significant at the 0.40 level or better.[8] The coefficient estimates are the incremental consumer surplus per unit of the variable. The function can be adjusted to predict consumer surplus for a trail according to specific characteristics of the trail. Consumer surplus measures per trail are reported in Table 12. These measures are calculated from the meta-analysis function by turning on (=1) or turning off (=0) the full effect of the trail-specific ($TRAIL_i$) and forest type (PP) dummy variables and adjusting each of the other variables according to the measure of the characteristic for a specific trail. If we use the same baseline consumer surplus measures as the target values used in previous assessments, validity measures ($\%\Delta V_{ij}$) can be calculated. These percent difference measures range from –62 percent error to –2 percent error, with an average percent error of 20 percent (Table 12).

We can now test the hypothesis ([9] and [10]) that the meta-analysis benefit transfer function provides more accurate measures of consumer surplus than using traditional approaches (Table 9). Table 13 provides the results of one-tailed paired t-tests on the validity measures for the meta-analysis transfer (Table 12) versus each of the different modeling strategies (by row comparison) in Table 9. The results show that we can reject the null hypothesis ([9]) that the two approaches result in equivalent levels of accuracy in favor of the alternative hypothesis ([10]) that the meta-analysis transfer approach is more accurate than the traditional value transfer approach for three out of six comparisons at the 0.10 significance level or better (Table 13). The three models that cannot be rejected at this significance level are those models that incorporate trail-specific travel cost shift variables (Model C) or the forest-specific models that are based on regional data (Models D and F, and Models E and G). At a significance level of 0.16 or better, we can reject the null hypothesis of equal accuracy in favor of the alternative hypothesis that the meta-analysis transfer is more accurate than the traditional approach (Table 13). This indicates that more specific modeling strategies result in more accurate transfers than ballpark estimates of value.

Conclusions

This study investigated how differences in the physical attributes of recreation sites are related to errors associated with a benefit transfer process. In addition, a meta-analysis benefit transfer function was developed to statistically relate measures of physical attributes of recreation sites with their associated benefit estimates. This study does not provide definitive evidence regarding the methodology and outcomes described herein. It begins the process of scientifically investigating the accuracy of benefit transfers and how different factors, as identified in the literature, affect them. As Berglund, Magnussen, and Navrud (1995) note, there is a need for research that specifically targets benefit transfers and/or is specifically designed for future benefit transfer applications. This study was undertaken for both of these reasons.

A site correspondence model is developed for hiking trips to several trails in Colorado. This model suggests that we can identify the effects of differences in the physical attributes of recreation sites on differences in value measures. Taking the constraints of the current dataset into account, we identified that error in the benefit transfer process was sensitive to differences in the sample's characteristics and differences in physical attributes of the sites. That is, a hiking trail is not a hiking trail if there are differences between them, such as degree of difficulty, landscape attributes, etc. It is not necessarily that these are different goods to be modeled differently, but that there are differences across the range of this good. These differences must be captured in a broad model if the model is to perform reasonably well in benefit transfers.

The meta-analysis model shows how the benefit transfer process is improved with models that are more sensitive to differences in the physical attributes of similar recreation sites. We recognize the possibility that the measures may be converging on the target values because we are repeatedly sampling from the same database. However, the hypothesis tests illustrate how modeling techniques that account better for differences between the sites performed better in benefit transfers than models providing ballpark estimates of values.

Two obvious improvements on this analysis can be identified. First, limitations of the data can be improved. For example, several of the single-site models, from which baseline or target values are derived, potentially suffer from small numbers problems. More observations for each of the models should improve their quality. In addition, the lack of modeling substitute sites may affect inter-site comparisons. Second, a broader range of physical attributes could be measured, especially using Geographic Information Systems technology, providing an added spatial dimension (Eade and Moran 1996). This may provide us with models that are more accurate, and subsequently more defensible, in applications to benefit transfers.

This analysis needs to be repeated under different circumstances. It would be interesting to see if a general pattern in the effects of different site characteristics emerges. This pattern, if it existed, would have a tremendous effect on how and what kind of data is collected in non-market valuation surveys.

References

Adamowicz, W.L., J.J. Fletcher, and T. Graham-Tomasi. 1989. "Functional Form and the Statistical Properties of Welfare Measures." *American Journal of Agricultural Economics* 71(2):414-421.

Berglund, O., K. Magnussen, and S. Navrud. 1995. *Benefit Transfer: Testing for Accuracy and Reliability*. Discussion Paper #D-03/1995. Norway: Department of Economics and Social Sciences. 21p.

Boyle, K.J., and J.C. Bergstrom. 1992. "Benefit Transfer Studies: Myths, Pragmatism, and Idealism." *Water Resources Research* 28(3):657-663.

Brouwer, R. 2000. "Environmental Value Transfer: State of the Art and Future Prospects." *Ecological Economics* 32(1):137-152.

Brouwer, R., and F.A. Spaninks. 1999. "The Validity of Environmental Benefits Transfer: Further Empirical Testing." *Environmental and Resource Economics* 14:95-117.

Creel, M.D., and J.B. Loomis. 1990. "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72(2):434-445.

Desvousges, W.H., F.R. Johnson, and H.S. Banzhaf. 1998. *Environmental Policy Analysis with Limited Information: Principles and Applications of the Transfer Method.* Mass: Edward Elgar.

Desvousges, W.H., M.C. Naughton, and G.R. Parsons. 1992. "Benefit Transfer: Conceptual Problems in Estimating Water Quality Benefits Using Existing Studies." *Water Resources Research* 28(3):675-683.

Downing, M., and T. Ozuna Jr. 1996. "Testing the Reliability of the Benefit Function Transfer Approach." *Journal of Environmental Economics and Management* 30(3):316-322.

Eade, J.D.O. and D. Moran. 1996. "Spatial Economic Valuation: Benefits Transfer using Geographical Information Systems." *Journal of Environmental Management* 48(2):97-110.

Englin, J., 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):104-112.

Englin, J., and T.A. Cameron. 1996. "Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7(2):133-147.

Feather, P., and D. Hellerstein. 1997. "Calibrating Benefit Function Transfer to Assess the Conservation Reserve Program." *American Journal of Agricultural Economics* 79(1):151-162.

Fletcher, J.J., W.L. Adamowicz, and T. Graham-Tomasi. 1990. "The Travel Cost Model of Recreation Demand: Theoretical and Empirical Issues." *Leisure Sciences* 12:119-147.

Kirchhoff, S. 1998. *Benefit Function Transfer vs. Meta-Analysis as Policy-Making Tools: A Comparison.* Paper presented at the workshop on Meta-Analysis and Benefit Transfer: State of the Art and Prospects, Tinbergen Institute, Amsterdam, April 6-7, 1998.

Kirchhoff, S., B.G. Colby, and J.T. LaFrance. 1997. "Evaluating the Performance of Benefit Transfer: An Empirical Inquiry." *Journal of Environmental Economics and Management* 33(1):75-93.

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

Loomis, J.B., A. Gonzalez-Caban, and J. Englin. 2000. *Testing for Differential Effects of Forest Fires on Hiking and Mountain Biking Demand and Benefits.* Fort Collins, CO: Department of Agricultural and Resource Economics, Colorado State University. Working Paper.

Loomis, J.B., B. Roach, F. Ward, and R. Ready. 1995. "Testing the Transferability of Recreation Demand Models Across Regions: A Study of Corps of Engineers Reservoirs." *Water Resources Research* 31(3):721-730.

Parsons, G.R., and M.J. Kealy. 1994. "Benefits Transfer in a Random Utility Model of Recreation." *Water Resources Research* 30(8):2477-2484.

Rosenberger, R.S., and J.B. Loomis. 1999. "The Value of Ranch Open Space to Tourists: Combining Observed and Contingent Behavior Data." *Growth and Change* 30(3):366-383.

-----. 2000a. "Panel Stratification in Meta-Analysis of Economic Studies: An Investigation of Its Effects in the Recreation Valuation Literature." *Journal of Agricultural and Applied Economics* 32(3):459-470.

-----. 2000b. "Using Meta-Analysis for Benefit Transfer: In-Sample Convergent Validity Tests of an Outdoor Recreation Database." *Water Resources Research* 36(4):1097-1107.

-----. 2001. *Benefit Transfer of Outdoor Recreation Use Values: A Technical Document Supporting the Forest Service Strategic Plan (2000 Revision).* General Technical Report RMRS-GTR-xxx. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 59 p. In Press.

-----. Forthcoming. "Benefit Transfer." In P. Champ, K. Boyle, and T. Brown, eds., *A Primer on Non-Market Valuation.* Under contract with Kluwer Academic Publishers, 2001.

Sturtevant, L.A., F.R. Johnson, and W.H. Desvousges. 1998. *A Meta-Analysis of Recreational Fishing.* Durham, NC: Triangle Economic Research. Unpublished Paper.

VandenBerg, T.P., G.L. Poe, and J.R. Powell. 2001. "Assessing the Accuracy of Benefits Transfers: Evidence from a Multi-Site Contingent Valuation Study of Groundwater Quality." In J.C. Bergstrom, K.J. Boyle, and G.L. Poe, eds., *The Economic Valuation of Water Quality.* Mass: Edward Elgar.

---

[1] Other information transfers can include using information on price and demand elasticities, on the magnitude and variance of price and quantity change measures, on economic multipliers and expenditure patterns, or adapting dose-response functions.

[2] This statement implies that primary research is conducted properly. In cases of poor survey design, benefit transfer may provide more accurate measures of the parameters of interest.

[3] That is, in some cases where primary research is possible and warranted, funds may not be allocated as an investment in the stock of this public good.

[4] Our stock of knowledge may depreciate because of a variety of factors, including evolving political and cultural climates or advances in valuation methodology.

[5] Research that comes the closest is research that tested inter-regional transfers, which implies differences between sites (Downing and Ozuna 1996; Loomis et al. 1995; VandenBerg, Poe and Powell 2001).

[6] N-1 refers to the approach wherein data is pooled for all but one trail (Loomis 1992). This results in a ballpark estimate of the value to be transferred to the n^{th} site.
[7] V_{Pj} is itself an approximation of the unknown but 'true' value of hiking at a specific site. Because this value is unknown, there is an error associated with it. For our present purposes, V_{Pj} as estimated is the assumed 'true' value for each trail. This is the traditional approach used when testing the validity of benefit transfers.
[8] *Trail 3* also measures the effect of *WATER*, since they are perfectly correlated dummy variable specifications.

Table1 Summary of Benefit Transfer Validity Tests

Study	Resource/Activity	Transfer Approach Category & Error (%)	
		Value	Function
Loomis (1992)	Fishing	5 – 40	5 – 15
Parsons and Kealy (1994)	Water quality	4 – 34	1 – 75
Loomis et al. (1995)	Recreation	---	1 – 475
Berglund et al. (1995)	Water quality	25 – 45	18 – 41
Downing and Ozuna (1996)	Fishing	---	1 – 34
Kirchhoff et al. (1997)	Rafting	24 – 56	6 – 228
Kirchhoff (1998)	Recreation/Habitat	---	1 – 7028
Brouwer and Spaninks (1999)	Biodiversity	27 – 36	22 – 40
Rosenberger and Loomis (2000a)	Recreation	---	0 – 319
VandenBergh et al. (2000)	Water quality	1 - 239	0 - 298

Adapted from and expanded on Brouwer (2000).

Table 2 Description of Variables

Variable	Description
TC	Travel cost as individual's reported share of transportation costs plus time costs (1/3 wage rate * travel time) (dollars)
$TC*F_i$	Interaction terms of TC and National Forest (i=3) (dollars)
$TC*T_i$	Interaction terms of TC and Trail (i=6) (dollars)
GENDER	Dummy variable; 1 = male, 0 = female
AGE	Respondent's age (years)
EDU	Respondent's level of education (years completed)
INC	Gross annual household income of respondent (dollars)
ELEV	Trailhead elevation above sea-level (feet)
GAIN	Elevation gain of trail from trailhead to summit (feet)
LONG	Length of trail (miles)
FIRE	The negative of the age of a wildfire in the recreation area (years)
WATER	Dummy variable; 1 = presence of water (lake, stream) near trail, 0 = no water
CROWN	Dummy variable; 1 = extreme fire in past as evidenced by crown fire, 0 = otherwise
PP	Dummy variable; 1 = presence of ponderosa pine trees, 0 = otherwise
LP	Dummy variable; 1 = presence of lodgepole pine trees, 0 = otherwise
ASPEN	Dummy variable; 1 = presence of aspen groves, 0 = otherwise

Table 3 National Forest and Trail Summary Statistics

National Forest/ Trail	N[a]	ELEV	GAIN	LONG	FIRE	WATER	PP	LP	ASPEN
Arapahoe/Roosevelt (Forest 1)									
Mount Margaret (Trail 1)	13	7800	100	5	-50	0	1	0	0
Grey Rock (Trail 2)	52	5400	2055	6	-8	0	1	0	0
Kilpecker/Blue Lake (Trail 3)	10	9400	1450	10	-42	1	0	1	0
Pike/San Isabel (Forest 2)									
Devil's Lookout (Trail 4)	25	8900	600	3	0	0	1	0	1
Gunnison/Uncompaghre (Forest 3)									
North Bank/Doc Park (Trail 5)	6	8600	900	7	-2	0	1	1	1
Summerville/DoubleTop (Trail 6)	21	8900	1400	9	-50	0	0	1	0

[a]n = number of respondents per trail.

Table 4 Summary Statistics of Sample (n=127)

Variable	Mean	Standard Deviation	Range
GENDER	0.51	0.50	0 – 1
AGE	36.48	11.22	19 – 73
EDU	16.12	2.22	11 – 20
INC	$68760	45326	5000 – 175000

Table 5 Multi-Site Travel Cost Models: Negative Binomial Random Effects

Variable	State Multi-Site			Forest Multi-Site[d]			
				Forest 1		Forest 3	
	A	B[a]	C[b]	D	E[c]	F	G[c]
Constant	26.5068	3.5402[e]	4.5289[e]	5.4956[e]	1.3207	1.2353	1.3450
	(23.95)	(1.25)	(1.42)	(1.49)	(0.95)	(1.26)	(1.50)
TC	-0.0122[e]	-0.0130	-0.0153[e]	-0.0153	-0.0446	-0.0275[e]	-0.0309
	(0.00)	(0.01)	(0.00)	(0.02)	(0.03)	(0.01)	(0.03)
TC*F1	---	-0.0089	---	---	---	---	---
		(0.02)					
TC*F2	---	-0.0033	---	---	---	---	---
		(0.02)					
TC*T1	---	---	0.0074	---	---	---	---
			(0.03)				
TC*T2	---	---	0.0039	---	0.0232	---	---
			(0.02)		(0.04)		
TC*T3	---	---	-0.1392[e]	---	-0.0911	---	---
			(0.05)		(0.07)		
TC*T5	---	---	-0.0300	---	---	---	---
			(0.04)				
TC*T6	---	---	-0.0090	---	---	---	0.0042
			(0.02)				(0.03)
GENDER	0.4921[e]	0.4478[e]	0.5130[e]	0.7880[e]	0.9262[e]	-0.3280	-0.3289
	(0.21)	(0.24)	(0.23)	(0.27)	(0.26)	(0.49)	(0.49)
AGE	---	---	---	0.0081	0.0074	-0.0440[e]	-0.0442[e]
				(0.01)	(0.01)	(0.03)	(0.03)
EDU	-0.1320[e]	0.0976[e]	-0.1497[e]	-0.2002[e]	---	---	---
	(0.05)	(0.05)	(0.06)	(0.08)			
INC	--	---	---	---	---	-0.0000	-0.0000
						(0.00)	(0.00)
ELEV	-0.0020	---	---	---	---	---	---
	(0.002)						
GAIN	-0.0034	-0.0004[e]	-0.0004[e]	---	---	0.0030[e]	0.0028[e]
	(0.003)	(0.00)	(0.00)			(0.00)	(0.00)
LONG	0.0215	0.0370	0.0086	-0.2842[e]	-0.1111	-0.1920[e]	-0.1859[e]
	(0.06)	(0.04)	(0.05)	(0.11)	(0.15)	(0.10)	(0.10)
FIRE	0.0173	---	---	-0.0040	-0.0136	---	---
	(0.02)			(0.01)	(0.01)		
WATER	-0.6894	---	---	---	---	---	---
	(1.32)						
CROWN	---	0.1937	0.2344	---	---	---	---
		(0.38)	(0.42)				
PP	-5.3073	-0.4249	-0.6369	---	---	---	---
	(5.97)	(0.66)	(0.92)				
LP	0.4055	-0.9030	-0.3797	---	---	---	---
	(0.64)	(0.66)	(0.91)				

ASPEN	1.6070	---	---	---	---	---	---
	(4.02)						
Alpha	1.3237[e]	1.5992[e]	1.5144[e]	1.5173[e]	1.5733[e]	0.6619[e]	0.6622[e]
	(0.18)	(0.21)	(0.20)	(0.27)	(0.29)	(0.24)	(0.24)
Adj.-R^2	0.08	0.03	0.02	0.12	0.10	0.14	0.12
Log-Lik	-530.43	-547.89	-542.24	-333.32	-335.62	-103.58	-103.56
# obs	254	254	254	150	150	54	54

[a]Forest 3 (Gunnison-Uncompaghre National Forest) is the omitted forest dummy variable. TC in Model B is the forest-specific travel cost variable for forest 3.
[b]Trail 4 (Devil's Lookout trail in the Pike-San Isabel National Forest) is the omitted trail dummy variable. TC in Model C is the trail-specific travel cost variable for trail 4.
[c]Trail 1 (Mount Margaret trail) in Model E and trail 5 (North Bank/Doc Park trail) in Model G are the omitted trail dummy variables for forest 1 (Arapaho-Roosevelt National Forest) and forest 3 (Gunnison-Uncompaghre National Forest), respectively. TC in each of these models is the trail-specific travel cost variable for the omitted trail in each model, respectively.
[d]Forest 2 (Pike-San Isabel National Forest) does not have a multi-site travel cost model because there is only one trail included in this forest.
[e]Variable is significant at the 0.10 level or better. Standard errors reported in parentheses.

Table 6 N-1 Multi-Site Travel Cost Models: Negative Binomial Random Effects

Variable	Model (Omitted Trail)					
	H (Trail 1)	**I (Trail 2)**	**J (Trail 3)**	**K (Trail 4)**	**L (Trail 5)**	**M (Trail 6)**
Constant	1.7004[a]	2.7159[a]	3.2836[a]	4.1673[a]	3.2958[a]	3.4316[a]
	(0.67)	(1.07)	(0.79)	(0.92)	(0.79)	(1.02)
TC	-0.0086[a]	-0.0104[a]	-0.0137[a]	-0.0314[a]	-0.0135[a]	-0.0079
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
GENDER	0.2477	0.4585	0.5496[a]	0.4916[a]	0.5430[a]	0.7747[a]
	(0.21)	(0.35)	(0.22)	(0.23)	(0.22)	(0.22)
EDU	-0.0806[a]	-0.1119[a]	-0.1502[a]	-0.1370[a]	-0.1495[a]	-0.1669[a]
	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
GAIN	0.0003[a]	-0.0006[a]	0.0001	-0.0002	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LONG	-0.0130	-0.0112	-0.0412	-0.0589	-0.0317	-0.1000
	(0.04)	(0.06)	(0.05)	(0.05)	(0.49)	(0.08)
FIRE	-0.0081	-0.0245[a]	-0.0189[a]	-0.0090	-0.0179[a]	-0.0209[a]
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
WATER	-1.4246[a]	-1.4757[a]	---	-1.8462[a]	-1.9403[a]	---
	(0.47)	(0.61)		(0.57)	(0.55)	
Alpha	1.1360[a]	1.4822[a]	1.4701[a]	1.3934[a]	1.4483[a]	1.5092[a]
	(0.16)	(0.27)	(0.20)	(0.20)	(0.20)	(0.22)
Adj.-R^2	0.02	0.07	0.06	0.08	0.05	0.06
Log-Like	-454.07	-304.10	-511.56	-442.88	-513.19	-451.06
# Obs.	228	150	234	204	242	212

[a]Variable is significant at the 0.10 level or better. Standard errors reported in parentheses.

Table 7 Trail-Specific Travel Cost Models: Negative Binomial Random Effects

Variable	Trail 1	Trail 2	Trail 3	Trail 4	Trail 5	Trail 6
Constant	3.3161	3.4528[a]	0.4365	2.1632	1.0793[a]	0.8109
	(3.00)	(1.17)	(0.41)	(1.70)	(0.52)	(1.58)
TC	-0.0825[a]	-0.0166	-0.0240	-0.0040	-0.0296	-0.0215[a]
	(0.04)	(0.01)	(0.05)	(0.01)	(0.03)	(0.01)
GENDER	1.9451[a]	0.5043[a]	---	0.4669	---	-0.9334[a]
	(0.78)	(0.27)		(0.49)		(0.52)
EDU	-0.0858	-0.1498[a]	---	-0.0278	---	0.0568
	(0.18)	(0.08)		(0.10)		(0.08)
AGE	---	---	---	-0.0480[a]	---	---
				(0.02)		
INC	---	---	---	0.0000	---	---
				(0.00)		
Alpha	1.4647[a]	1.1272[a]	0.4257	0.5319[a]	0.8059	0.8216[a]
	(0.18)	(0.24)	(0.68)	(0.26)	(0.87)	(0.29)
Adj.-R^2	0.26	0.04	0.01	0.11	0.09	0.12
Log-Like	-67.99	-226.10	-25.05	-79.43	-23.07	-83.73
# Obs.	26	104	20	50	12	42

[a]Variable is significant at the 0.10 level or better. Standard errors reported in parentheses.

Table 8 Baseline and Transfer Consumer Surplus Measures per Trip (Equation 1)

			Baseline Measures (V_{Pj})			
Model	**Trail 1**	**Trail 2**	**Trail 3**	**Trail 4**	**Trail 5**	**Trail 6**
Trail	$12.12	$60.38	$41.56	$248.85	$33.81	$46.57

			Transfer Measures (V_{Ti})			
Model	**Trail 1**	**Trail 2**	**Trail 3**	**Trail 4**	**Trail 5**	**Trail 6**
A	$81.93	$81.93	$81.93	$81.93	$81.93	$81.93
B	45.56	45.56	45.56	61.29	76.91	76.91
C	126.14	87.62	6.47	65.45	22.09	41.21
D	65.50	65.50	65.50	---	---	---
E	22.43	46.73	7.37	---	---	---
F	---	---	---	---	36.41	36.41
G	---	---	---	---	32.37	37.40
H	116.42	---	---	---	---	---
I	---	95.76	---	---	---	---
J	---	---	73.01	---	---	---
K	---	---	---	31.83	---	---
L	---	---	---	---	73.81	---
M	---	---	---	---	---	126.82

Table 9 Benefit Transfer Validity Measures Using Traditional Value Transfer Approach

Model	Trail 1	Trail2	Trail 3	Trail 4	Trail 5	Trail 6	Average $\lvert\%\Delta V_{ij}\rvert$
A	575.99%[a]	35.69%	97.14%	-67.08%	142.32%	75.93%	165.69%
B	275.91	-24.54	9.62	-75.37	127.48	65.15	96.35
C	940.76	45.11	-84.43	-73.70	-34.66	-11.51	198.36
D & F	440.43	8.48	57.60	---	7.69	-21.82	107.20
E & G	85.07	-22.61	-82.27	---	-4.26	-19.69	42.78
H – M	860.56	58.60	75.67	-87.21	118.31	172.32	228.78

[a]Validity (percent difference) measures = $[(V_{Ti} - V_{Pj})/V_{Pj}]*100$ (equation 6).

Table 10 Site Correspondence Model: Random Effects

Variable	*Coefficient*	**Standard Error**	**Significance Level**
Constant	87.2429	121.34	0.47
%ΔGENDER	45.4459	11.93	0.00
%ΔAGE	22.4562	14.58	0.12
%ΔELEV	-9.7922	5.72	0.09
%ΔGAIN	-0.7929	0.32	0.01
%ΔLONG	12.3915	5.80	0.03
%ΔFIRE	-0.0284	0.14	0.84
%ΔWATER	-4.6403	1.92	0.02
%ΔPP	0.2824	3.75	0.94
%ΔLP	-2.1720	2.41	0.37
%ΔASP	10.4403	8.12	0.20
Adj-R^2	0.79		
# Obs.	34		

Dependent variable is $\%\Delta V_{ij}$, or validity (percent difference) measures (Table 9).

Table 11 Meta-Analysis Benefit Transfer Function: Random Effects

Variable	Coefficient	Standard Error	Mean of Variable
Constant	-2602.9263	2704.68	---
TRAIL 1	7.9950	11.15	0.175
TRAIL 3	-20.6873[a]	11.17	0.175
TRAIL 4	-5.2568	12.52	0.125
TRAIL 5	-9.9429	12.20	0.175
TRAIL 6	2.5347	12.04	0.175
ELEV	0.2381	0.20	7514.542
GAIN	0.5439	0.47	1341.859
LONG	-75.8477[a]	26.73	6.332
FIRE	-8.9823	10.25	-22.661
PP	618.0245	729.57	0.700
Adj. $-R^2$	0.72		
# Obs.	40		

Dependent variable is consumer surplus per trip (table 8).
[a]Variable is significant at the 0.10 level or better.

Table 12 Meta-Analysis Consumer Surplus Estimates and Validity Measures

Trail	V_{Ti}	V_{Pj} [a]	$\%\Delta V_{ij}$
Trail 1	$4.54	$12.12	-62.54%
Trail 2	56.88	60.38	-5.80
Trail 3	35.16	41.56	-15.40
Trail 4	242.90	248.85	-2.39
Trail 5	29.36	33.81	-13.16
Trail 6	36.82	46.57	-20.94
Avg. $\lvert\%\Delta V_{ij}\rvert$			20.04

[a]From trail-specific models, see table 8.

Table 13 Results of Hypothesis Tests

Model Comparisons	t-Stat	Significance Level
Meta vs. Model A	1.95	0.05
Meta vs. Model B	2.36	0.03
Meta vs. Model C	1.26	0.13
Meta vs. Models D, F	1.13	0.16
Meta vs. Models E, G	1.45	0.11
Meta vs. Models H – M	1.76	0.07

Hypothesis Tested: H_0: $\%\Delta V_{METAij} = \%\Delta_{TRADij}$ vs. H_1: $\%\Delta V_{METAij} < \%\Delta_{TRADij}$.

Experiments in Valuing Wetland Ecosystems

John P. Hoehn,

Frank Lupi,

and

Michael Kaplowitz

Michigan State University

Presented at the Annual Meeting of Western Regional Project W-133
Miami Florida
February 26-28, 2001

Contact: John P. Hoehn, Department of Agricultural Economics, Michigan State University, East Lansing, MI 48864; Ph: 517-353-6735; Fx: 517-432-1800; E-mail: hoehn@msu.edu. (Web presentation available at: http://www.msu.edu/user/hoehn. Additional research information at: http://wetlands.rd.msu.edu).

Abstract: A utility-theoretic model indicates that mitigation prices for wetland ecosystems depend on preferences *and* technical knowledge. Empirical analysis found gaps in respondents= knowledge about such ecosystems. Valuing wetland types requires dealing with respondents= possible misinformation, by developing tools for informing respondents or by combining service-based valuations with valid technical data.

Wetlands ecosystems are valued for a range of ecological services. These services are protected by national, state, and local regulation. The primary federal wetland protection statute is Section 404 of the Clean Water Act (33 U.S.C. '1344). Under this statute, the U.S. Army Corps of Engineers, in conjunction with the U.S. Environmental Protection Agency (EPA), administers a review and permitting process for the Adischarge of fill material@ in Awaters of the United States.@ Since 1989, the guiding principle of federal wetland policy is the Ano net loss@ of wetlands criterion (Gaddie and Regens 2000). To implement this principle, the wetland permit process encourages potential dischargers to minimize and avoid wetland impacts wherever possible. Where wetlands are impaired or destroyed, wetland mitigation is required.

Mitigation refers to actions taken to recreate, restore, or protect wetlands of an equivalent type and function to those being impaired or destroyed (Denison and Schmid 1997). Since wetlands vary by type, ecological functions, and the services they yield to humans, the means for judging the equivalency of destroyed and mitigated wetlands is both problematic and central to successful implementation of the Ano net loss@ policy (National Research Council (U.S.). Committee on Characterization of Wetlands. 1995; Mitsch and Gosselink 1993). Substantial effort has been made to define and measure wetland equivalencies using engineering principles and biophysical characteristics (Bartoldus 1999). However, the economic equivalency of wetland services has received less attention. Absent an understanding of the economic tradeoffs, wetland mitigation may leave economically important services unprotected and under provided.

In this paper, we report initial research results regarding the development and application of a framework for measuring the relative economic values of wetland ecosystems. We begin by reviewing the ecological characteristics of wetland ecosystems and past efforts to value wetlands. We then derive a model that leads to three approaches to estimating wetland ecosystem values in stated choice experiments. The relative performance of these valuation approaches depends on the distribution and extent of ecological knowledge among respondents. Knowledge of a particular form is an essential input into accurate ecosystem valuation.

The second part of the paper examines the knowledge base that residents of central Michigan might use in valuing wetland ecosystems. Residents were contacted using random digit dialing and were asked to participate in a group discussion about natural resource issues. Each group involved 6 to 8 residents. Each group interview was conducted by a moderator using a prepared discussion guide.

Discussion participants demonstrated better than expected general knowledge of wetland ecosystems, but their detailed knowledge of wetland functions and services was uneven. Participants recognized habitat for plants and animals as a key wetland function. A smaller portion identified maintenance of water quality and water storage as important wetland functions. Misperceptions were also revealed. For example, several respondents thought that trees do not grow in wetlands and that wetlands kill trees despite the fact that wooded wetlands are common in Michigan. When asked to interpret and discuss photographs of wooded wetlands, these participants said that wetlands were killing the trees.

Wetlands Ecosystems and Valuation Research

Wetlands are transitional types of ecosystems that occupy a spectrum between land and water ecosystems. Their exact definition has been controversial (National Research Council (U.S.). Committee on Characterization of Wetlands. 1995). The operational definition used in Federal wetlands permitting regulations builds on two essential wetland characteristics: (i) the land is composed of soils that are water-saturated during part of the vegetation growing season and (ii) the land supports plants that are typical of saturated soils (Smith et al. 1995). Using this definition, wetlands may have covered about 12 percent of the area of the continental United States during colonial times. Since that time, human activity in the United States has converted approximately 45 percent of wetlands area to other uses (Heimlich, Carey, and Brazee 1989).

Wetlands ecosystems vary greatly in type, ecological function, and services to human beings. Wetland types include bottomland swamps, tidal marshes, cattail marshes, vernal ponds, fens, and bogs. Ecological functions of wetlands include water storage, maintenance of surface and groundwater flows, biochemical cycling, retention of water-suspended and dissolved materials, accumulation of peat, maintenance of characteristic biological energy flows, and maintenance of characteristic habitats.

Wetland types and functions provide services that affect human well-being. The water storage function, for instance, may result in service to human beings by retaining floodwaters. Maintenance of groundwater flows may contribute to stable sources of potable water. Wetland habitats may offer recreational opportunities, open space amenities in otherwise densely settled areas, and potential non-use services such as maintaining biodiversity.

The objective of wetland mitigation is to replace wetlands destroyed by permitted activities through the creation, restoration, or protection of equivalent wetlands. The ratio of mitigated wetland area to impaired wetland area is called the mitigation ratio. Mitigation ratios typically vary by wetland type. For instance, in Michigan, recent rules require compensatory mitigation of 1.5 acres for each acre lost when the wetland being lost is a common type. When the destroyed acreage is a rare wetland type, 5 acres of mitigation are required for each acre lost (MCL '324.30319). At the Federal level, the Army Corps of Engineers makes adjustments in the mitigation ratios to account for the type and duration of impacts, the rarity of the impacted wetlands, and the methods used in mitigation (U.S. Army Corps of Engineers-Charleston District 1996).

Wetland mitigation ratios are analogous to the in-kind prices of impaired wetlands. Such ratios represent an agency=s in-kind valuation of mitigation activities relative to the lost wetland type or function. A question then arises regarding the adequacy of such prices. For instance, a mitigation ratio that is satisfactory on engineering or biological grounds, may not be acceptable in terms of preventing the loss of economic services and values. For instance, a particular wetland may be ecologically common in a region or state, but rare in terms of its recreational services and open space amenities by virtue of its location in an urban area. Hence, using Michigan=s rules to make the point, the statutory mitigation ratio for replacement of a particular cattail marsh might be set at 1.5 to 1 on statewide ecological grounds, whereas the particular wetland=s economic value to its urban area might warrant a rare wetland ratio of 5 to 1.

The economic literature suggests the importance of considering relative economic values in mitigation pricing. Many studies estimate the value of specific wetlands and thereby demonstrate the economic value of wetlands. However, most studies shed little light on the relative value of different wetlands types, functions, and wetland services (Heimlich et al. 1998). A handful of studies do document commercial and recreational values associated with some wetlands (Loomis et al. 2000; Costanza et al. 1998; Bergstrom and Stoll 1993).

Rr research suggests that wetlands may provide open space amenities (Mahan, Polasky, and Adams 2000; Opaluch 2000). Some recent studies imply that the economic services of wetlands, including recreation, water quality, and flood control services are well recognized by ordinary citizens (Azevedo, Herriges, and Kling 2000). Especially interestingly in terms of mitigation ratios, Mullarkey (1997) estimates that an acre of naturally occurring wetland is 6 times more valuable to respondents than an acre of mitigated wetland.

Woodward and Wui (2001) conducted a meta-analysis to estimate the value of wetland services. The estimated values of services per wetland acre are shown in Figure 1. The estimates indicate that consumptive services such as bird hunting (Hunt Bird), commercial fishing (Com Fish), and recreational fishing (Rec Fish) created significant derived demand for wetland protection. Non-consumptive services were also highly valued (Habitat, Flood Cntrl, and Bird Watch). Bird watching (Bird Watch), for instance, had the highest value per acre of all the wetland services examined in the study.

Economic Features of Wetland Ecosystems

Wetlands mitigation, to varying degrees in different cases, attempts to account for differences in wetland types, functions, and services. In the context of mitigation, economic values are useful to the extent that they allow for differences across wetland ecosystem types, functions, and services. In an economic sense, a wetland is not a generic economic commodity. Rather, a wetland is a Lancastrian, multi-attribute bundle that may vary in three major dimensions: type, function, and service. A research design for wetland ecosystem valuation would vary these attributes and assess how value changes with changes in ecosystem type, function, or service.

A second feature of wetland ecosystems that bears on the economics of wetland values is that wetland attributes occur in specific patterns and types. Ecosystems share a general pattern of species relationships. At the foundation of an ecosystem food web are plants that convert energy and nutrients into food. Plant consumers and predator relationships are build upon the vegetative foundation. The specific pattern of species relationship varies with the type and scale of an ecosystem (Miller 1999). That is, a fen does not support the same species and relationships as a bog. Nor does a small wetland of a particular type support the higher order predators that a larger wetland of the same might (Osborn 1996). Since the species mix and interrelationships may vary with type and scale, it is possible that the economic value of wetland types may differ from individually valued sets of wetland functions and services.

A third feature of wetland ecosystems that impacts the economics of wetland values is the uncertainty associated with incomplete knowledge. Knowledge of wetland ecosystems, their functions and services is incomplete on the scientific level (Miller 1999). That is, science may not be able to characterize a full list of relevant wetland attributes nor may science be able to help restore these attributes once there are impaired. In turn, ordinary citizens have incomplete and possibly inconsistent knowledge of the science of wetland ecosystems and functions. Given the evolving nature of science, a useful economic research design for ecosystem valuation might describe how wetland values change with specific changes in respondents= baseline knowledge of wetland types, functions, and services.

A Research Design for Wetland Ecosystem Valuation

The research design outlined below takes an initial step toward a rigorous framework for valuing wetland types and services in stated preference experiments. The goal is a research design that shows the relationship between the value of wetlands as wetland types and the value of wetland as Lancasterian service bundles. As our research program advances, we plan to extend the framework to describe the derived demand for wetland functions. Additionally, we seek a wetland valuation design that makes explicit the role of respondents= knowledge in valuation.

The framework that addresses two wetland types. Wetland acreage of type 1 is represented by $A1$. Wetland acreage of type 2 is represented by $A2$. Each wetland type yields different sets of wetland services. Wetland type 1 yields services of a single kind that we represent with the symbol $S1$. Wetland type 2 yields services of the first kind, $S1$, as well as services of a second kind, $S2$. The total amounts of services available from acreages of type 1 and 2 are:

$$(1) \qquad S1 = A1 + A2 \qquad S2 = K(A2)$$

where K(A2) is an increasing, concave function that maps the acreage of type 2 into a levels of services S2. Equation (1) might correspond to a situation where both wetlands provide open space amenities but only type 2 wetlands support habitat with significant biodiversity.

The next step in the valuation model is to link economic services with human well-being. Human well-being is represented by a utility function, U,

$$(2) \qquad U = U(S1,S2,M),$$

where the level of well-being depends on the levels of the two services and an economic measure of income, M. The link between wetland acreage and well-being comes from the combination of equations (1) and (2). Substituting equations (1) into (2) shows the relationship between economic well-being and wetland acreages,

$$(3) \qquad \begin{aligned} U &= U[A1 + A2,\, K(A2),M] \\ &= u[A1,A2,M] \end{aligned}$$

where u() is utility function defined on wetland acreage rather than services. This latter utility function leaves the relationship between acreage and services implicit.

In economic terms, a no-net loss policy would leave economic well-being unchanged by compensating for a reduction in type 2 acreage with an increase in type 1 acreage and visa versa. For small changes in acreage, the amount of compensatory mitigation required to offset the loss of type 2 acreage is derived by taking the total differential of the second line of equation (3) with respect to U, $A1$, and $A2$. To keep well-being constant, dU is set equal to zero and the differentials rearranged. By this method, the following economic mitigation ratio is derived,

$$(4) \qquad P_{A2\,A1} = Mu/MA2\ /\ Mu/MA1$$

$P_{A2\,A1}$ is the utility-theoretic mitigation price of a small reduction in type 2 acreage, measured in terms of a compensating increase in type 1 acreage. In terms of the utility function, this mitigation price is the ratio of the marginal utility of type 2 acreage, M*U/MA2*, and the marginal utility of type 1 acreage, M*U/MA1*.

Each of the marginal utilities in equation (4) is potentially measurable in stated choice experiments. In a choice experiment, respondents would be presented with alternative policy choices involving wetland acreage of type 1 and type 2. The choice data for acreage could then be used to statistically estimate the marginal utilities. Similar experiments could be conducted for choices involving wetland services such as open space and biodiversity. The problem then becomes how to link the estimated marginal utilities of services to the mitigation choices characterized in terms of acreage.

The link between the mitigation price of acreage, $P_{A2\,A1}$, and the mitigation price for services of type 1 and 2, $P_{S2\,S1}$, may be derived by taking the total differential of the first line of equation (3) with respect to *U*, *S1*, and *S2*. Setting *dU* equal to zero leads

$$
\begin{aligned}
(4) \qquad P_{A2\,A1} &= 1 + K{=}\ Mu/MS2\ /\ Mu/MS1 \\
&= 1 + K{=}\ P_{S2\,S1} \\
&\$\ 1
\end{aligned}
$$

where *K=* is the marginal productivity of acreage of type 2 in producing services of kind 2, as understood and known by choice experiment respondents.

Several features of the mitigation price as stated in equation (4) are notable. First, we can expect the mitigation price of acreage to be greater than one when the in-kind price of services is positive. Mitigation with a wetland type that offers fewer services than the wetland type lost requires an acreage premium. Thus, the mitigation ratio between a wetland that is more diverse in services and one that is less diverse in services is greater than one.

Second, the mitigation price is a function of preferences as represented by the marginal utilities and by the perceived technical relationship between acreage and the second kind of service. This technical relationship is represented by K= in equation (4). The marginal utilities of acreage estimated in stated preference experiments are conditioned on respondents= knowledge of K=. If respondents= knowledge is inconsistent with wetland science, the mitigation prices may be inconsistent with wetland science as well.

Respondents= knowledge plays a central role in accurate estimation of the marginal utilities of acreage. If this knowledge is inconsistent with wetland science, there seem to be two ways to bring the mitigation prices in line with the science. First, it may be possible to bring respondents= knowledge in line with scientific knowledge using educational tools such as carefully worded text, photographs, and diagrams. Whether such informational devices can be effective is an open hypothesis that warrants appropriate tests.

A second way to bring mitigation prices in line with the science is to design stated preference experiments to elicit the mitigation price of services, $P_{S2\,S1}$. The wetland service preference information, $P_{S2\,S1}$, may be combined with a scientific estimate of K= to calculate a facsimile acreage mitigation price based on scientific information,

$$
(4) \qquad S_{A2\,A1} = 1 + k{=}\ P_{S2\,S1}
$$

where k= is the scientific measure of the marginal productivity of type 2 acreage in producing services of the type 2 kind.

The analysis of the economic model of ecosystem values leads to three alterative valuation approaches shown in Table 1. Each approach varies in its information requirements regarding individuals= preferences and the ecological relationship between acreage and services. One approach sets up the choice experiments in terms of acreage tradeoffs for different wetland types. Such an approach mixes preference with ecological knowledge in the structure of the mitigation prices. All else equal, it results in a valid estimate of mitigation prices if respondents= knowledge is adequately complete and consistent with science.

The second approach sets up the wetland ecosystem choice experiments in terms of tradeoffs in ecosystem services. Such an approach would compliment the preference information from respondents with information on ecological relationships from science. It would yield a mitigation price based on science that the researcher deems appropriate and acceptable. The science portion of the valuation may also be modified as scientific information changes. A drawback to this approach is that the list of relevant services identified by the research and specified in the model may be incomplete resulting in a partial valuation. In addition, such an approach may not capture the value associated with the pattern of ecological relationships represented by wetland types.

A third approach to wetland ecosystem valuation is based on wetland types. This approach modifies the first approach by attempting to bring respondents= knowledge in line with scientific knowledge. This approach would try to assess respondents= baseline knowledge and to develop information tools that would alter the baseline so that respondent=s knowledge was consistent with scientific knowledge. Respondents would engage in choice experiments once they received a systematic exposure to the information treatment. A key issue for the success of this method is whether respondents are sufficiently sensitive to the new information. If not, the new information may have little effect and the choice experiment results would mirror those of the first approach.

The availability of three different approaches to valuing wetland ecosystems offers the opportunity for cross-corroboration and hypothesis testing. For instance, the second approach based on scientific information might be used to set reasonable upper bounds on the valuation estimates derived from the first approach. Further, the second approach might be used to set up hypotheses regarding the effects of information treatments on the mitigation price.

Knowledge Base of Michigan Respondents

Qualitative research is helping us learn what it is that people value about wetland ecosystems. This step will be used to help the researchers determine the functions and services that should be the focus of the valuation effort. Furthermore, the qualitative research also gives insights into the general state of people=s knowledge about wetland ecosystems, their functions, and types (Kaplowitz 2000). We have also been exploring ways of communicating to respondents about wetland functions, Awhat wetlands do.@

To this point, the qualitative research has conducted three group discussions with participants recruited from the general population of adults in the Lansing, Michigan. Each discussion group involved 6 to 8 participants. Participants were initially contacted using random selected telephone numbers. Because of election year resistance to participate in political focus groups, participants were asked to participate in a group discussion of Anatural resource issues in Michigan.@ They were not told that we would be discussing wetlands.

Outline of group interviews

Each group interview lasted for roughly two hours. Sessions were held in a facility on the campus of Michigan State University. All of the sessions were conducted by the same moderator who used the same discussion guide for each session. The moderator used non-directive prompts to encourage participants to participate and elaborate their responses. The discussion guide and the sessions had five basic sections, with the first three taking roughly 45 minutes and the last two sections taking roughly 45 minutes. The balance of the time was used for breaking the ice, taking a Asnack@ break, or completing university paperwork.

The five substantive sections of the discussion guide and sessions were:

1. Introduce participants, identify each participant=s top three natural resource issues, and discuss.
2. General background questions about wetlands to explore what participants know about wetlands and to learn about their experiences with wetlands and the things that wetlands do.
3. Photographs of both wetland and non-wetland ecosystems projected on a screen to determine how people judge what is and is not a wetland, to see if people can distinguish wetland and non-wetland plant communities, and to see if people know about different types of wetlands.
4. Verbal, written, and graphic presentation of different wetland functions including flood control, wildlife habitat, and sediment retention. The functions and definitions for this section were taken from scientific literature on wetlands.
5. Some questions about wetland mitigation and about replacement of impaired wetlands. In the later two focus groups, there were additional questions about replacing wetlands lost due to a highway project were used.

Knowledge of wetland functions

Participants evidenced knowledge of wildlife habitat functions of wetlands. The participants also rated the wildlife habitat functions highly in terms of their relative importance vis-à-vis other wetland ecosystem functions. Almost all participants rated wildlife habitat as extremely important, the highest category, on their function ranking worksheets. This finding is consistent with other research on wetlands (Azevedo, Herriges, and Kling 2000; Swallow et al. 1998; Stevens, Benin, and Larson 1995).

Participants had mixed knowledge of some of the other functions of wetland ecosystems such as water quality, groundwater recharge and flood control. Often there were a few respondents in each focus group that were aware of and knowledgeable about one or more of these Anon-habitat@ functions. However, every group had a majority of participants who seemed much less aware of these types of functions and who did not seem very knowledgeable about them.

Interestingly, several of the scientifically recognized wetland functions prompted negative feedback from participants. Several individuals rejected the importance of functions such as pollution interception and waste treatment. These individuals expressed strong opinions that wetlands should not be used for these functions. In several instances, participants voiced their concern that environmental laws are supposed to provide for pollution cleanup and waste treatment; wetlands need not perform such functions. Note that these functions appear prominently in much of the literature describing wetland functions. After further discussions, most of these participants felt that it would be all right to create new wetlands for purposes such as waste treatment. This feedback seems to illustrate the potential difficulty of relying solely on scientific descriptions of wetland ecosystems, functions, and services.

What do photographs communicate?

As a part of the group sessions, photographs of various wetlands were shown to the participants. This exercise was intended to probe participants= knowledge of wetland types, wetland vegetation, and general understanding of wetland ecosystems. The participants= discussions of the images yielded some interesting insights about what photographs can communicate to people. For example, at one point we showed a photograph of a fen (a particular wetland type) that did not have visible water and had grasses and vegetation that was browning. In response to this image, some respondents noted that it did not look healthy and that it was not supposed to be that way. One participant said the photograph showed an area that AI would say [was] scorched by fire.@ In reality, the photograph contained a moderate amount of shadow that was mistaken as evidence of fire. This photograph clearly communicated something other than what had been intended, and the cue that caused the misperception, the shadows, is unlikely to be absent in future photographs of fens and other ecosystems.

Another example of the power of photographs to (mis)communicate was found when the blurry background in a photograph of a non-wetland meadow was Aseen@ by a respondent to be water. It is important to note that the focus group participants were viewing these images on a large projection screen at levels of resolution that are likely quite higher than what would be feasible in a typical survey application. The conclusion that can be drawn from these experiences is that photographs do communicate information, both intended and unintended, and that they must be pre-tested along with other potential survey elements. This will hold for web-based surveys as well as other mediums.

Wetland mis-perceptions

As a part of the group interviews, participants were shown a variety of photographs that depicted different wetland types in different settings as well as photographs that did not show wetlands. Part of the group interview probed for whether or not each of the photographs depicted a wetland. In each of the groups, several respondents commented on the notion that trees do not grow in wetlands and that wetlands kill trees. In fact, some participants used their perceived presence of dead trees in the photographs to distinguish wetlands from non-wetlands. Therein lies the source of the paper title. The so-called Adead tree@ comments occurred in all three of the sessions and they occurred in relation to different photographs of forested wetland areas. It is interesting to point out that in Michigan where the participants live over two-thirds of the wetlands are forested.

Another factor that may have played a role in this perception was that one of the wetland photographs showed some prominent trees that had been attacked by Dutch Elm disease. However, two of the sessions raised comments about wetlands and dead trees in conjunction with photographs of forested wetlands shown before the image of the wetland with the diseased trees. Thus, the photograph with the dead elms did not cause the perception, though it may have amplified the perception for some individuals. One conclusion that emerges from these examples is that it seems vital to the design of an accurate valuation instrument that researchers be aware of respondents= perceptions (and mis-perceptions) about the good being valued. Establishing such information is a key step in the development of methods of communicating with respondents about the good to be valued and the context of the valuation.

Understanding of mitigation

In all three of the group sessions, some questions were asked to about wetland mitigation and about the replacement of impaired wetlands. These questions were aimed at revealing peoples= understanding and acceptance of wetland mitigation. In the later two sessions, additional questions were asked in the context of a scenario in which the government would be replacing wetlands impaired by a highway project. This scenario was developed to force people to consider, to add realism, and reinforce the idea of trade-offs. The scenario was also used to learn more about one possible context for stated preference wetland valuation. The comments and discussion surrounding these portions of the group sessions revealed a general skepticism that wetland mitigation could adequately replace what might be lost due to a wetland impairment. This skepticism is related to the unique challenges posed by ecosystems as well as the role of knowledge as an input into ecosystem valuation.

Another finding from this section of the group interviews was that there was some confusion over the meaning of wetland mitigation, especially wetland replacement. Some individuals took the concept quite literally and inferred that it would mean transferring plants and animals from one site to the mitigation site. For example, one participant asked, AHow are they going to transfer all those frogs?@ Again, this serves as another example of how indispensable to survey design it is for researchers to have a grasp of respondents= baseline knowledge and understanding.

Perhaps the main finding from what was learned about peoples= knowledge of mitigation relates to the general skepticism about replacing all functions of a specific wetland. The following are examples of the kinds of comments we received in discussions on wetland mitigation:

AI don=t know if you can come out equal.@

AReally replacing or just duplicating parts you see?@

ALike substituting oleo for butter.@

ACould they truly get back all that was lost?@

It appears that such skepticism consists of two elements. The first related to a disbelief that certain functions, or services, of wetlands could actually be replaced. The second related to a feeling by several individuals that wetland replacement would not adequately compensate for impairments because wetlands are complex. That is people acknowledge that even though many functions might be replaced, there is more to the wetland than the specific functions that get replaced. Both elements of peoples= skepticism raise issues that are fundamental to ecosystem valuation. The former element raises questions about whether we want to elicit people=s beliefs in the underlying production relationship, K(.), at the same time we elicit economic choices and values. As illustrated above in the table, this can lead to a co-mingling of values and knowledge about how final services are derived from the Areplacement@ wetland ecosystem. The second element speaks to the notion that an ecosystem is more than a bundle of listed functions or services.

Conclusions

The valuation framework outlined above identifies three approaches to valuing wetland ecosystems and wetlands mitigation. The three approaches show that the economic value of wetlands is derived from the value of wetland services; wetlands are valued when they yield valuable services. This linkage between wetlands and wetland services has an important implication for stated choice experiments. If respondents= knowledge is inconsistent with wetland science, stated choice experiments may yield incomplete or inaccurate valuations.

Knowledge of the linkage between wetlands and wetland services plays a slightly different role in each of the three valuation approaches derived above. The first valuation approach takes respondents= knowledge as given. It elicits a valuation conditioned on respondents= baseline knowledge. The second approach elicits a valuation of wetland services and then uses scientific knowledge to compute a wetland valuation from the estimated value of services. The third approach attempts to bring respondents= knowledge in line with scientific knowledge using systematic information treatments. It elicits wetlands values conditioned on respondents= updated knowledge base.

The reported qualitative research was intended to explore the knowledge base of likely respondents in order to assess the feasibility of the three valuation approaches. Initial findings show that Michigan residents are more cognizant of wetlands than expected, but that their knowledge is uneven. Most respondents had some prior knowledge of wetlands functions such as provision of wildlife habitat, maintenance of groundwater flows, and floodwater retention. However, some functions identified by wetland science, such as retention of polluted run-off and waste treatment, were rejected as illegitimate by some respondents. A portion of these respondents thought that pollution retention would harm the ability of a wetland to support wildlife and other functions. Others thought that current environmental laws should lead to cleanup of pollution at the source, rather than letting pollution flow into a wetland.

The qualitative research also underscored the difficulties of using photographs to communicate wetland knowledge. The initial hypothesis was that photographs might be an effective means of communicating differences in wetlands types and functions. Photographs, however, seemed to be an inaccurate communication device. When shown a photograph of a fen, some respondents correctly interpreted dark areas as shadows, while others interpreted the same dark areas as evidence of impairment and, perhaps, fire. When shown photographs of wooded wetlands, some respondents concluded that the wetlands were killing the trees, even though healthy wooded wetlands are a common wetland type in Michigan.

The evidence thus far underscores the role of knowledge as an input in valuing wetland ecosystems. The empirical results show that respondents have some baseline knowledge of wetlands, but that this baseline knowledge may be incomplete or inaccurate in certain dimensions. In this context, each of the three valuations approaches may be useful in posing and testing hypotheses about wetlands values and the effect of knowledge. For instance, if respondents= baseline knowledge is incomplete, values estimated via the second approach may be larger than values estimated via the first approach. Thus, the three valuation approaches may offer the means of testing and corroborating wetland values.

References

Azevedo, Christopher, Joseph A. Herriges, and Catherine L. Kling. 2000. Iowa wetlands: Perceptions and values: Center for Agricultural and Rural Development.

Bartoldus, Candy C. 1999. *A comprehensive review of wetland assessment procedures: A guide for wetland practitioners*. St. Michaels, MD: Environmental Concern Inc.

Bergstrom, John C., and John R. Stoll. 1993. Value estimator models for wetlands-based recreational use values. *Land Economics* 69 (2):132-137.

Costanza, Robert, Ralph d'Arge, Rudolf de Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg, Shahid Naeem, Robert V. O'Neill, Jose Paruelo, Robert G. Raskin, Paul Sutton, and Marjan van den Belt. 1998. The value of the world's ecosystem services and natural capital. *Ecological Economics* 25:3-15.

Denison, Mark S., and James A. Schmid. 1997. *Wetland mitigation: Mitigation banking and other strategies for development and compliance*. Rockville, MD: Government Institutes Inc.

Gaddie, Ronald K., and James L. Regens. 2000. *Regulating wetlands protection*. Edited by M. E. Kraft and H. Wilson, *Suny series in environmental politics and policy*. Albany: State University of New York Press.

Heimlich, R.E., M.B. Carey, and R.J. Brazee. 1989. Beyond swampbuster: A permanent wetland reserve. *Journal of Water Conservation* 44 (5):445-450.

Heimlich, Ralph E., Keith D. Wiebe, Roger Claassen, Dwight Gadsby, and Robert M. House. 1998. Wetlands and agriculture: Private interests and public benefits. Washington, D.C.: Resource Economics Divisions, Economic Research Service, U.S. Department of Agriculture.

Kaplowitz, Michael D. 2000. Identifying ecosystem services using multiple methods: Lessons from the mangrove wetlands of Yucatan, Mexico. Agriculture and Human Values 17 (2):169-179.

Loomis, John B., Paula Kent, Liz Strange, Kurt Fausch, and Alan Covich. 2000. Measuring the total economic value of restoring ecosystem services in an impaired river basin: Results from a contingent valuation survey,. *Ecological Economics* 33 (1):103-117.

Mahan, Brent L., Stephen Polasky, and Richard Adams. 2000. Valuing urban wetlands: A property price approach. *Land Economics* 76 (1):100-113.

Miller, G. Tyler. 1999. *Environmental science: Working with the earth*. Belmont, CA: Wadsworth Publishing Company.

Mitsch, William J., and James G. Gosselink. 1993. *Wetlands*. 2nd ed. New York: Van Nostrand Reinhold.

Mullarkey, Daniel J. 1997. Contingent valuation of wetlands: Testing sensitivity to scope. Dissertation, Agricultural and Applied Economics, University of Wisconsin - Madison, Madison, WI.

National Research Council (U.S.). Committee on Characterization of Wetlands. 1995. *Wetlands: Characteristics and boundaries*. Washington, D.C.: National Academy Press.

Opaluch, James. 2000. Strength of preference indicators. Paper read at EPA Environmental Policy and Economics Workshop Series, October 2000, at Washington, D.C.

Osborn, Ronald G.. 1996. Ecological criteria for evaluating wetlands. Fort Collins, CO: National Biological Service.

Smith, Daniel, Alan Ammann, Candy Bartoldus, and Mark Brinson. 1995. An approach for assessing wetland functions using hydrogeomorphic classification, reference wetlands and functional indices. Washington, D.C: U.S. Army Corps of Engineers.

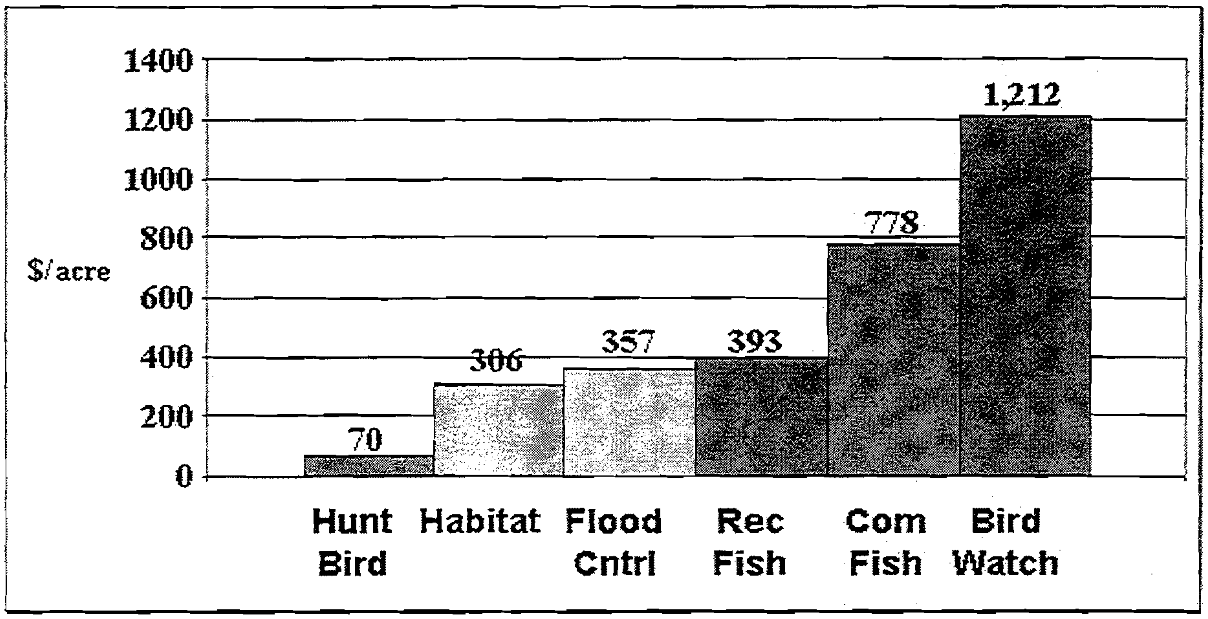
Stevens, Thomas H., Samuel Benin, and Joseph S. Larson. 1995. Public attitudes and economic values for wetland preservation in new england. *Wetlands* 15:226-231.

Swallow, S., M. Spencer, C. Miller, P. Paton, R. Deegen, L. Whinstanley, and J. Shogren. 1998. Methods and applications for ecosystem valuation: A collage. In *Proceedings of the first workshop in the environmental policy and economics workshop series*, edited by B. Kanninen. Washington, D.C.: Office of Research and Development and Office of Policy, U.S. Environmental Protection Agency.

U.S. Army Corps of Engineers-Charleston District. 2000. *Compensatory mitigation*. 1996 [cite available from http://www.sac.usace.army.mil/permits/sop96-01.pdf.]

Woodward, Richard T., and Yong-Suhk Wui, AThe Economic Value of Wetland Services: A Meta-Analysis,@ Department of Agricultural Economics, Texas A & M University, College Station, TX, 2001.

Figure 1. Value of Wetland Services



Woodward and Wui, 2001

Table 1. Valuation Approaches

	Choice Experiment Design	Limitations
1	Tradeoffs in terms of acreage of different wetland types	Confounds preferences and ecological knowledge; Biased if respondents= knowledge is incomplete or inconsistent
2	Tradeoffs in terms of final wetland services	Incomplete service list; miss value of whole
3	Tradeoffs by acreage type, but make systematic effort to provide scientific information	Perceptions may not be sensitive to scientific information

Does Money Matter in Assessing Global Warming Policies? Price Signals, Green Fervor, and Lifestyle Costs

S.L. Johnson

and

P.R. Wandschneider

Presented at the Annual Meeting of the Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Abstract: In this study, we asked about 180 students in general education environmental science classes about their attitudes toward global warming issues and their valuation of four alternative levels of carbon controls. On a series of attitudinal questions we found that students generally showed concern for global warming and favored forceful policies to reduce the potential effects of global warming – at least as strong as those in the Kyoto Accords. A modified contingent ranking approach was used to estimate values for the four policy levels. The four levels of carbon emission controls were: status quo, 7% below 1990 levels (Kyoto accord levels for the USA), 25% reduction, and 50% reduction. We asked respondents to compare each alternative with the others for a total of 6, two-way choices at different prices. We used a total of 10 pricing schemes, in all of which a policy that had greater emission reductions was higher priced than one with lesser reduction. From these two-way choices we constructed rankings for the four alternatives at different relative prices. We used a simple multinomial logit model on the unranked and ranked data to analyze the results looking at just price as an independent variable. We also constructed an expanded multinomial logit model including responses to some demographic, income and attitudinal questions as explanatory variables.

Introduction

The potentially catastrophic effects of global warming make it a particularly visible and important public issue. For instance, it was the cover story of the April 9, 2001 issue of TIME magazine. This cover was part of a spate of articles which emerged following the release of a new report by IIPC which shows a consensus of scientists predicting global warming effects and attributing at least some of the cause to human activities. This report proposed a wide range of global warming effects, reflecting the lack of consensus among research scientists about the probable effects of global warming. Despite the increasing case being made for the potential impacts of global warming, no consensus has emerged about either the extent or type of action that should be taken, at least in the United States. In April 2001, the President of the United States indicated the U.S. will not sign the international agreement known as the Kyoto Protocol as Congress and policy pundits debate the wisdom of different policy options. Part of the reason that policy is uncertain is that the causes of global warming and the nature and extent of its potential impact are uncertain. Another part of the policy quandary concerns how to make policy under the very long time lags that exist between actions and effect. Still, current policy choices, whether by commission or omission, affect the future state of the world. Can individual citizens meaningfully evaluate these policy choices? It is important to increase our understanding of citizen attitudes and values towards potential global warming policies: what are people willing to do now to ameliorate potential future damages? It is particularly interesting to consider the attitudes and values of young adults. Young adults will have to live with the eventual results of any current policy decisions. Moreover, they are entering their age of majority, when they will have responsibility to make the decisions and to pay for them – both monetarily and in terms of their opportunities and experiences in the world in which they will live. In this study, we report results from a study of the attitudes and values of this critical population of young adults.

Economic critics of contingent valuation and some environmentalists share a distrust of using survey data to place economic values on environmental attributes. This study makes an interesting test ground for some of the issues that divide those who believe that contingent valuation is a sound basis for estimating economic values and those who believe the method has limited, if any, value. In this specific case, both the sample population and the target policy appear to be questionable subjects for a CVM study. Students are not "real" households. They are in a transitory period of their life, with uncertain future prospects for income, occupation, and other important dimensions of their future. Moreover, it might be more difficult to formulate clear, well-formed opinions about a policy topic (global warming) that is replete with long time lags and large measures of uncertainty. We also note, with interest, that some environmental activists argue that <u>money should not matter</u> for environmental policies. People should approve the "right thing" regardless of cost. We suggest that, if evidence of distinct willingness-to-pay can be found in this population and for this topic, it suggests the CV method should be much more robust for more stable respondent groups and better-defined questions. That is, we think this is a case of testing a method when it is potentially most likely to fail.

In this study, we asked about 180 students in general education environmental science classes about their attitudes toward global warming issues and their valuation of four levels of carbon emission controls. We used a modified contingent ranking approach to rank these four alternatives at different relative prices. We found that students generally showed concern for global warming and favored aggressive policies to reduce the potential effects of global warming – at least as strong as those in the Kyoto Protocol. The willingness to pay variable was clearly significant and appropriately signed for all policies: increases in the price reduced the probability that students would approve the policy. The inferred willingness to pay was high. Additionally, two interesting features emerge from the data that may have wider significance for how contingent valuation studies are done. The first feature of the data is a problem that emerges in defining a relevant and meaningful income category for students. Formation of a meaningful demand requires some income reference point. If students are thinking about prospective policies, what is the relevant income standard: their current budget seems arbitrary; their parents' income is relevant but unclear; and their future income is largely unknown. In this paper we describe the basic study and develop and discuss results -- especially results pertaining to the issue of defining a student's income and to possible explanations for the apparent anomaly in values for different policy types.

The second interesting feature is that students' willingness to pay for a moderate policy is greater than their willingness to pay for a stricter policy (more emission reductions). Students appear to be willing to pay more (money) for less (carbon reduction) – a result seemingly at odds with basic economic logic. This apparent anomaly has different implications depending on the interpretation of its meaning. One interpretation is that one simply should not expect strong results from a study using respondents who have relatively unformed preferences over a topic as diffuse as global warming. This interpretation implies that there is nothing to be learned from this study for the general body of contingent valuation work or for global warming policy. We will argue that the current study must certainly be classified as a pilot study since it used a convenience sample, but the sample size, the method, and the results are hold up as indicative and interesting.

A second explanation is that price is an unimportant part of this sample groups' evaluation of alternative global warming policies. Fervent "greens" support strong policies, moderates support moderate policies, and non-greens support minimalist global warming policies. Respondents are simply expressing the strength of their attitudes and price and willingness to pay have little to do with their responses. In the view of some authors (Kahneman and Knetsch, 1992; Diamond and Hausman, 1993) many contingent valuation studies, especially of "existence" or passive use values merely show attitudes, not valid willingness to pay. In this interpretation then, a finding that a moderate policy has the greatest appeal regardless of price supports the idea that this contingent valuation study does not really measure willingness to pay. We will argue that there is evidence for an effect from "green fervor" in our data, but that ALSO there are clear indications that students are responding in a meaningful way to the willingness to pay question. We conclude that the simple dichotomy between those who view CV as "working" and those who view CV as meaningless is not just simple -- it is too simplistic. We argue that results from the current study indicate that the issue is more one of calibrating the CV derived values than it is of either rejecting or accepting them wholesale.

Additional explanations for the value anomaly arise from the possibility that respondents may impute attributes to policies beyond those clearly articulated in the study instructions. The whole point of contingent valuation is that values are contingent on specification of the policy package. For instance, consider the probability of success of a given policy. In our description of policies we implicitly presume that a policy targeted to a 25% reduction in CO_2, will achieve its objective. Respondents may not share this belief. In fact, twenty-eight percent were unsure of the effectiveness of any government program to ameliorate global warming. Another very intriguing result concerns the possibility of non-monetary, life-style costs of a policy. Suppose that one policy would cost $X per year and another would cost $Y per year AND require that one spent more time commuting. This later problem is akin to the problem that arises if the time component is omitted in travel cost studies. Our results suggest that these additional policy dimensions may partly account for the apparent inconsistency in preferences. We think the possibility of a "lifestyle" cost dimension is something that may occur in many policy evaluation scenarios. Our results suggest that studies should attempt explicit incorporation of such costs in the policy scenario definitions.

Survey

The survey instrument comprised a paper and pencil questionnaire administered to students in two general education environmental studies classes at Washington State University. The survey was voluntary and anonymous, although one must assume that there was some degree of pressure to participate since it was a class activity. The survey took about 20 minutes to complete. Teaching assistants administered the test in lab sections and they were trained to provide identical instructions to all sub-groups. There were 181 responses. The classes were environmental studies classes for non-majors. Business and education were the two most well represented majors. To clarify – these respondents were definitely not environmental science majors, although their choice of environmental science to satisfy a science general university requirement may imply some relevant self-selection. Students had been recently exposed to material about global warming in the class. The respondents appeared to be fairly representative of a broad range of first and second year non-science majors at Washington State University - that is, they were mostly white, middle-class young adults (study details and demographics available from authors).

The survey instrument had four basic sections – the fourth section being a standard socio-demographics section eliciting social and demographic information about the respondent, such as age and gender. The first section of the survey was devoted to determining familiarity with and concern about the environment. The survey was relatively rich in information and questions about attitudes toward global warming. It began with some general environmental attitude questions. A short definition of global warming was followed by an optional one-page description of global warming. (We asked them to indicate if they read the section.)

The second section asks questions about Global Warming and carbon dioxide abatement. It contains questions concerning respondents' beliefs about global warming, followed by a short description of the Kyoto agreement and a discussion of the potential for reducing those impacts through reductions in CO_2 levels (see appendix table). It then asks how much the respondents support the Kyoto agreement.

The survey next moves to questions about the respondents' willingness to pay for different levels of carbon dioxide abatement. The value elicitation questions were framed as a set of trade-offs among four alternative policies labeled K (7% below 1990 levels of emissions – the USA target from the Kyoto Protocol), B (25% reduction from 1990 levels), C (50% reduction), and SQ (status quo). The levels were chosen to be relevant, clear and distinct. The payment vehicle was gas and energy taxes. Each respondent ranked all four alternatives as paired alternatives. Prices were varied according to a scheme that always included the status quo at zero price, and wherein each succeeding increment in carbon reduction had a higher price. Altogether, there were 10 policy price schemes. Prices were based around actual cost estimates contained in Cline (1992). The price scheme and the number of respondents who received each scheme are shown in table 1. Since the SQ option is always included at no cost, it is not shown in the table. Note that the bottom two rows show the (weighted by N) average and median of the prices elicited.

Table 1: Range and Averages of Policy Costs per Person ($/year)

Version	7%	25%	50%	N
A	15	35	50	16
B	15	50	220	21
C	35	50	110	17
D	35	75	350	16
E	50	75	350	19
F	50	110	750	18
G	75	110	350	18
H	75	220	2000	20
J	110	220	750	19
K	220	750	2000	17
Averages	**7%**	**25%**	**50%**	
Total mean	67.62	167.82	691.27	
Total median	50.00	110.00	350.00	

Attitudes

As noted above, we elicited responses to a large number of attitudinal questions. Many of the questions used were adopted from other national and international environmental research (for example, Dunlop et al, 1992; Dunlop and Scarce, 1991). While a detailed discussion of the attitudinal results is not within the scope of the present paper, some highlights provide a useful context for the value questions. Twenty-three percent of respondents believe that environmentalists exaggerate the dangers of global warming, but three-quarters believe that experts agree that global warming is real. Also three-quarters agree that global warming is human caused. Only one-quarter said that they would only approve efforts by the United States to reduce CO_2 emissions if all other nations also reduced emissions. Finally, three quarters (76%) believe that the United States should ratify the Kyoto Protocol. Some specific attitudinal and belief questions were used in the statistical analysis and these are identified in later parts of the paper.

CVM Analysis

The four policy levels were arranged into 6 two-way comparison questions. *Sample language: "For each pair, circle the number to indicate which policy you prefer: Q-12. Prefer 1. policy C: 50% reduction; $50/year 2. policy SQ: no action."* Responses to the six questions were compiled to produce an ordering of all four policies for each respondent at the price scheme used for that respondent's survey. Table 2 shows the raw first choice rankings (at the bid values given in the surveys). Incomplete and intransitive rankings were discarded from further analysis. Only 3 of the 181 respondents returned surveys with intransitive rankings. These raw rankings reveal that the status quo was rarely the first choice policy under the costs given to respondents in the survey.

Table 2: Raw Percentages: Most Preferred Policy	
Policy K most preferred by	23 %
Policy B most preferred by	31 %
Policy C most preferred by	25 %
Policy SQ most preferred by	4 %
Incomplete or intransitive answers	17%

Table 3: Raw Percentages: Paired Preferences	
Policy K preferred to Policy SQ by	81 %
Policy B preferred to Policy SQ by	75 %
Policy C preferred to Policy SQ by	54 %
Policy K preferred to Policy B by	28 %
Policy K preferred to Policy C by	51 %
Policy B preferred to Policy C by	51 %

Because the responses are qualitative, we used multinomial logit to analyze the data (LIMDEP software). The analysis used 150 of the 181 responses after the incomplete and intransitive responses were discarded. Logit and the closely related probit methods are used to estimate the likelihood that one alternative will be chosen from other available choices. Multinomial logit generalizes conventional logit -- simple logit is employed when the dependent variable is a single qualitative variable such as a simple choice with a yes-no response. Logit and probit models are the most commonly used techniques for modeling qualitative dependent variables – where logit and probit differ only in the assumed distribution of the error term. The distribution of the error turns out to make the logit (lognormal distribution) model easier to estimate than the probit (normal distribution) model. We first estimated models using unranked and ranked dependent variable data with one explanatory variable. Unranked multinomial models estimate the probability that one answer is chosen over all alternatives (the three other choices in this case). Ranked multinomial uses the ranking information. Finally, we investigated the results in more fully specified, multi-variate models that included socio-demographic and attitudinal variables. For more detailed discussion of these methods, see, for example, Greene's econometric text.

First, consider the simple models with only price used as an explanatory variable. Table 4 shows results from the model using the unranked data. The multiple qualitative variables are each assigned a separate coefficient.

Table 4: Results from Multinomial Logit (unranked data)

Variable	Coefficient	t-value
cost ($/year)	-0.00082018	-2.462
Policy K Dummy	1.834363	4.473
Policy B Dummy	2.614058	5.549
Policy C Dummy	2.416257	5.266
Restricted log-likelihood: -183.9637		
Maximum log-likelihood: -180.4723		
Likelihood ratio test: 6.9828		

Results show that price (the cost of the policy in the questionnaire) had a clear and significant effect on the likelihood of respondent approval of any alternative. However, the intercept coefficient for each policy is different and highly significant (the status quo policy is the omitted alternative, so all coefficients are relative to policy SQ). Intriguingly, the intercept coefficient for policy B (25% reduction) is highest. This result implies that policy B is most preferred when the model controls for the price effect. The next table shows the marginal effects from costs on the likelihood of approval for different policies – that is, what happens to the probability of choice of each alternative when there is an increase in the cost of the preferred policy. (Marginal effects in a logit model are not directly revealed in the coefficients of the estimated equation but must be calculated.) Results show very clearly that an increase in price will reduce the likelihood that a respondent will choose that policy alternative and increase the probability that they will choose one of the other alternatives. There is a clear and economically logical price effect – but it is also relatively small in magnitude – demand is very inelastic. A $100 increase in price brings less than a 2% drop in probability of choice.

Table 5: Marginal Effects from Cost on Policy Choice (from unranked Multinomial Logit)

	Marginal effect of change in cost when :			
	Policy K is preferred	Policy B is preferred	Policy C is preferred	Policy SQ is preferred
Policy K(7)	**-0.016**	0.009	0.006	0.001
Policy B(25)	0.009	**-0.019**	0.009	0.001
Policy C(50)	0.006	0.009	**-0.017**	0.001
Policy SQ(0)	0.001	0.001	0.001	**-0.004**

Next we turn to the estimates using ranked data. Table 6 shows results from a multinomial logit model in which choice is determined by two sets of characteristics - the price of the policy alternative and its level of carbon reduction. This model permits extraction of marginal willingness to pay estimates by conventional techniques (Lareau and Rae, 1989).

Table 6: Results from Ranked Multinomial Logit (simple model)

Variable	Coefficient	t-value
Cost ($/year)	-0.0011662	-5.555
Emissions reduction % below 1990 levels	0.032549	9.409
Restricted log-likelihood: -362.2981		
Maximum log-likelihood: -332.2349		
Likelihood ratio test: 60.1264		

The median willingness to pay from this equation is $29 for each 1% reduction in emissions. Table 7 shows the median willingness to pay calculated for the policy alternatives used in this study. The estimated willingness to pay is at the upper end of the prices used in the elicitation schemes[i]. The results indicate a very high willingness to pay for emissions reduction. Also, in this model the prices are nicely ordered: the value of policy C is the greatest. However, this result is an artifact of the forced linearity of the WTP in the simple model and does not hold up under closer scrutiny.

Table 7: WTP for Emissions Reduction	
Policy	WTP value
no change in reduction level (SQ)	$ 0
7% reduction in emissions (K)	$ 195.37
25% reduction in emissions (B)	$ 697.76
50% reduction in emissions (C)	$ 1395.52

Table 8 shows results for the multinomial logit using only price as an explanatory variable but incorporating the ranking information. Use of the ranking information does not change fundamental results but improves the efficiency of the estimates. These results are very close to those for the unranked data, except that all coefficient estimates are more significant and the log-likelihood ratio test indicates that the over all significance of the model is very high. (Log-likelihood ratios cannot be directly compared for relative significance unless the models are nested.) Again, the price effect is significant. Again policy B has the highest intercept – indicating that it would be the favored choice at equal prices. This result is depicted more clearly in table 9. Table 9 also indicates that the median WTP would be much higher in this model and that policy B would have the highest median WTP.

Table 8: Results from Ranked Multinomial Logit (second model)		
Variable	Coefficient	t-value
Cost ($/year)	-0.0010472	-4.434
Policy K Dummy	1.99054	13.752
Policy B Dummy	2.68325	17.980
Policy C Dummy	2.32003	11.115
Restricted log-likelihood: -281.2596		
Maximum log-likelihood: -272.8494		
Likelihood ratio test: 16.8204		

Table 9: Probability Policy Preferred to No Action (SQ) at Different Price Levels

Policy	$50	$110	$350	$750	$2000	$5000
K	**0.874**	0.867	0.835	0.769	0.474	0.0375
B	0.933	**0.929**	0.91	0.87	0.643	0.0723
C	0.906	0.901	**0.876**	0.823	0.556	0.0514

Table 10 shows results for a model that includes income and attitude and belief information. The model shows a good fit as indicated by the high log-likelihood ratio. The variables are interacted with the policy choices so that separate coefficients are given for each variable's effect on each policy. Again, the policy intercept coefficient is very strong and shows that policy B (25% reduction) is preferred -- other things equal.

Two income variables were included: parents' income and expected future income. We attempted to get data on student's current disposable income but this proved very difficult. Students have heterogeneous sources of support (scholarships, own work, family). They also have a variety of arrangements for paying their expenses that will affect disposable income (on campus versus "Greek" versus apartments). We did not find a satisfactory method of estimating comparable figures for disposable income so we have no variable for current budget.

The two income variables used produced some intriguing results. Parental income has the expected effect – the higher the income, the more likely a respondent is to approve any policy. However, only the coefficient on policy K approaches statistical significance. We found no obvious interpretation of this except to note that policy C shows no sensitivity to parental income at all. Support for policy C in particular, and all policies to some extent, does not vary much with income group – we will discuss this further below. On the other hand it will turn out that some of the attitudinal variables are very significant.

Coefficients on future income are very clearly significant. These coefficients show that greater expected future income reduces the probability of choosing all policy options (relative to SQ) and that its effect is similar for all policy options. It is not clear what story this variable is telling. Is a decrease in carbon dioxide an inferior good relative to future income? This does not seem likely. One possible explanation is that expected future income is a proxy for profession, and that those expecting higher income (example, business majors) are less "green" than those expecting lower income (example, teachers). Under this interpretation, the future income variable should not be treated as an income variable at all, but as a predicator variable for lifestyle and professional choices that might be less or more "green." This is part of the story about preference heterogeneity that we discuss below.

The results from the expanded model help in the analysis of the different hypotheses for explaining the seemingly strange result that a policy reducing carbon by 50% is worth less to respondents than a policy that would reduce carbon by 25%. Recall that we have three basic hypotheses: this reversal of an expected economic relationship suggests 1) that survey results measure attitudes -- but not true WTP values; 2) that respondents are heterogeneous in their preferences for emissions policy on the basis of some unobserved variable which might be called "greenness;" 3) and that the emissions reduction policies have implicit secondary attributes that explain the WTP patterns.

The results shown in table 10 provide some support for the hypotheses that this study does not successfully measure WTP of the respondents. These lines of evidence include: the finding that the strongest emissions reduction policy is not the preferred policy when cost is removed; the weakness of the income variables; and the low marginal value of the price coefficient (inelasticity of demand) and the "unreasonably" high value of the inferred WTP measure.

As in the simple models, the intercept coefficient (dummy) on policy B in the expanded model is the highest so that, other things equal, policy B would be the preferred alternative. In fact, in this more fully specified model, the Kyoto policy has a higher intercept than does policy C. This result implies that, ceterus parabus, policy B (25% reduction) is preferred to policy K (7% reduction), which is preferred to policy C (50% reduction). If this result holds, it is contrary to the standard economic behavioral model that "more is better." However, we think the results from the expanded model suggest some very interesting explanations. We find that respondents have heterogeneous preferences over the desired degree of carbon emission control and that there are implicit secondary attributes of the policies to which the respondents are reacting. To explain this interpretation we must look at some of the other model variables. But before we discuss the attitude and policy specification issues, let us complete the discussion of the price and economic variables.

Table 10: Results from Multinomial Logit (many variables)

Variable	Coefficient	t-value
Cost ($/year) for Policy K	-0.00270231	-0.797
Cost ($/year) for Policy B	-0.00243561	-2.671
Cost ($/year) for Policy C	-0.00112880	-3.919
Policy K Dummy	5.891295	3.184
Policy B Dummy	6.405180	3.867
Policy C Dummy	4.525538	2.607
Parents' Income (on K)	0.2231328	1.916
Parents' Income (on B)	0.1307957	1.228
Parents' Income (on C)	0.0680128	0.575
Expected Future Income (on K)	-0.4797686	-3.676
Expected Future Income (on B)	-0.5272473	-4.350
Expected Future Income (on C)	-0.4345825	-3.219
age (on K)	0.04668876	0.558
age (on B)	0.19649572	2.697
age (on C)	0.32053002	4.289
Rating of Seriousness of Global Warming (Q-7 of Survey) (on K)	-1.186816	-4.788
Rating of Seriousness of Global Warming (Q-7 of Survey) (on B)	-1.504063	-5.949
Rating of Seriousness of Global Warming (Q-7 of Survey) (on C)	-1.656662	-5.080
Rating of Willingness to Drive Less (Q-8 of Survey) (on K)	0.03622836	0.201
Rating of Willingness to Drive Less (Q-8 of Survey) (on B)	-0.4363057	-2.458
Rating of Willingness to Drive Less (Q-8 of Survey) (on C)	-0.9003588	-4.560
Rating of Threat to Future from Environmental Problems (Q 4.6 of Survey) (on K)	-0.4411888	-3.291
Rating of Threat to Future from Environmental Problems (Q 4.6 of Survey) (on B)	-0.6767624	-4.960
Rating of Threat to Future from Environmental Problems (Q 4.6 of Survey) (on C)	-0.7314931	-4.929
Restricted log-likelihood: -281.2596		
Maximum log-likelihood: -211.4664		
Likelihood ratio test: 139.5864		

A meaningful and useful estimate of WTP must produce a measure that is comparable to prices paid in markets for other goods and services. The income variable is therefore not only important in establishing a literal "budget constraint" but also in calibrating prices to a common metric. After all, the basic goal is to translate the unmeasured underlying utility metric (preference orderings) into a common accounting system so that values can be compared and even aggregated across individuals. If some respondents are using (internal) "dollars" while others are using (internal) "yen" or "lira," then comparisons or aggregations have no meaning because we do not know the "exchange rate." Thus, the fact that no satisfactory income variable is available for this respondent group is potentially troubling. However, on closer examination, the real issue may not be the income variable itself – but whether or not respondents are using the same units in expressing their willingness to pay, and how these units correspond to intended behavior and to other prices paid. If all respondents are using the same metric, but the units are not strongly linked to a clear income level, than the issue is simply one of calibrating the responses.

The implication of this point is that, instead of two alternatives (true WTP versus invalid WTP) there are three alternatives for interpreting the WTP measures. Are the WTP estimates measuring: 1) true WTP; 2) "merely attitudes;" or 3) some X which is a (common) scalar multiple or some other simple function of true WTP? Therefore the test of whether or not the WTP estimate is legitimate is not simply a test of whether the value is of the right magnitude, but of whether respondents understand the cost-price variable in the same way. Our interpretation of the results is that respondents are producing consistent and comparable value estimates of the policy options. We are uncertain what their metric is, but, whatever reference they are using, it seems to be common to them all. Perhaps they are thinking of the price of the carbon emissions policy in terms of trade-offs with common student purchases such as music disks or movie rentals rather than as some expenditure subtracted from a known income. Support for this conclusion comes from a pooling test of the price variable. In a test not reported in this paper, we found that respondents were quite heterogeneous in their preferences, but that *they responded to the price variable the same way, regardless of their preferences among the four alternative policies*. By inspection, one can see that the coefficients on price are very close for the three alternatives in the expanded model. The most parsimonious explanation of the common price coefficient is that respondents are using the same "units" of measurement (at least up to a linear transformation - cardinal measurement) and responding to the price in a similar way. In short, the price coefficient measures some signal of "real" willingness to pay – though the magnitudes may need calibration.

Another aspect of the price coefficient could be taken to support the case that the present study does not capture true willingness to pay. Although the marginal price effect is significant it is relatively weak. As one can see from Table 9, it takes a very large price to induce half the respondents to oppose any of the policy options. A consequence of this is that the estimates of willingness to pay for each alternative would be very large. Are the policies too price-inelastic and the inferred estimates of willingness to pay too large to be credible? To some extent this is a matter of judgment because there is no external criteria by which to judge the figures. Our own conclusion is that, in this study, we have found a signal of WTP but the WTP value may not be correctly calibrated. The fact that the price variable is so very well behaved is noteworthy. It is significant and "rational:" increasing the cost of a policy decreases the probability that respondents will choose the policy. And, while the WTP estimates are high, they are not outside the range of the plausible.

The literature on the external validity of CVM willingness to pay measures is extensive. (For a survey of some of the issues see Mitchell and Carson.) While some studies show that willingness to pay estimates from CVM studies can be quite close to those found in revealed preference studies, others find situations in which CVM derived values may be high. Thus, there is a discussion in the literature of a potential upward bias from "yea-saying" in discrete choice contingent valuation studies (e.g. Boyle, et al., 1998). In fact, the NOAA "blue ribbon" panel recommended using discrete choice elicitation format and then dividing the estimate by 2 (Arrow, et al., 1993). Of course, economic theory suggests the possibility that respondents may answer strategically in public good situations – asserting a willingness to pay that is deliberately higher than their true willingness to pay. Another hypothesis is that there is an anchoring effect (Boyle, et al., 1998). Still another conjecture is that respondents may unintentionally bias responses (generally upward) because they are only hypothesizing about how they would really act. There may be discord between their stated preferences and what would turn out to be their revealed preferences. This "hypothetical bias" can be especially problematical when respondents construe the situation to be a donation rather than a required payment. In fact, psychological studies indicate that people overestimate their own "goodness" -- though their estimates of what others will do come closer to their own actual behavior than their predictions of their own behavior (Epley and Dunning, 2000). In summary, it may be that responding students are "overestimating" what they will really be willing to pay especially since they have limited experience with incomes and budgets. Still, overestimation of one's behavior is not the same as giving a "stated preference" that has ***no*** relationship to behavior -- as some critics seem to claim about CVM estimates. Again, the issue becomes one of calibration rather than the simple dichotomy: accept the stated value as true or reject it.

One concern about our conclusions might be raised. The statistical analysis was based upon data from which responses with inconsistent rankings of the policies had been excluded. Therefore, it might be claimed that the results of economic rationality were "cooked" in the sense that economically irrational responses were excluded. However, only three such inconsistent responses were found. Moreover, even though the data were purged of inconsistent and intransitive preferences over the alternatives, the remaining data still permit "irrational" responses -- such as an insignificant or inverted price variable. In the event, we found a significant and economically rational price coefficient.

Let us now return to the issue of the putative fundamental "irrationality" of the basic preference ordering: B>K>C>SQ. The a priori "correct" preference order should be C>B>K>SQ (ranked from most to least effective in reducing global warming) -- cost and other characteristics being equal. Taking these differences at their face value, the issue is, do the differences in basic preferences reflect differences in the respondents' attitudes (greenness) or do they reflect perceived differences among the policies – on dimensions other than emission reduction. We find evidence of both.

Examination of the coefficients on some of the attitudinal variables clearly indicates that greenness affects policy preference. In our test of the price variable we had found that respondents were similar in how they viewed the price variable, but that their preferences were quite heterogeneous over the policy alternatives. We also found that the heterogeneity corresponded very little to standard socio-demographic variables like sex. (Age is significant, but there is very little variation over age in this population so the implication of the age variable is unclear.) So, given the finding of preference heterogeneity we included three attitude questions in the expanded model. With the exception of one question on one of the policies, all coefficients are significant and are in the correct direction. (More environmental responses have lower values, so a negative on the coefficient indicates that the more "non-green" the respondent, the less likely that the respondent will approve the policy – and the reverse – the more green, the more likely to approve.) The deciding point in the "green fervor" evidence is that, the relationship between the greenness attitudinal variables and the likelihood approval is stronger, the stronger the emission reduction policy. If one is more likely to think that global warming is serious, then one is even more likely to approve policy C than B and B than K – and all at very significant t-values. Similarly, as one's belief that environmental problems are a threat to the future increases, the likelihood of approving all the policies relative to SQ increases, but it increase more rapidly for C than B and more rapidly for B than K.

After "controlling for" green attitudes with these attitudinal variables, there remains an underlying preference for policy B which we think is due to perceived differences in attributes of the policies that were not explicitly identified in the survey. We hypothesize that two policy characteristics might influence attitudes toward the policy alternatives: likelihood of implementation and lifestyle cost. Since we did not explicitly include these dimensions in our policy descriptions, we must look to indirect evidence concerning their role in policy preferences. One variable in the extended model speaks to "lifestyle" costs. We asked respondents about their willingness to drive less to reduce the probability of global warming damage. The results show a very strong relationship between willingness to "sacrifice" by driving less, and approval of any of the non-status quo policies. Again, a revealing point is that the coefficient is greater for the stronger policies -- indicating that increased willingness to sacrifice is associated with a higher inclination to approve a stronger policy. Again the coefficients are well ordered in this dimension, with the coefficient effectively 0 on the Kyoto alternative, an intermediate coefficient and significance on the 25% (B) policy, and highest and most significant on the 50% (C) policy. The results for this variable suggest to us that willingness to make lifestyle changes is a perceived dimension of global warming policy, and that it affects preferences over the policies. In addition, recall that the negative correlation on future income (those with higher expected future incomes are less likely to approve carbon emissions reduction policies) might be interpreted as an indication that respondents perceive higher emission reduction policies as a threat to their standard of living.

The expanded model includes no direct variable on the relationship between the credibility of policies and the likelihood of policy approval. We found no clear relationship between policy preferences and responses to a question about whether the respondents thought that a government policy to reduce emissions would be likely to be effective.

Conclusions

In this study, we asked about 180 students about their attitudes toward global warming issues and their valuation of four levels of carbon emission controls. The willingness to pay variable was clearly significant and appropriately signed for all policies: increases in the price reduced the probability that students would approve the policy. The inferred willingness to pay was high.

The present study has two features that reduce the potential to generalize results for either policy or disciplinary purposes. First, the study uses a convenience sample of students. Besides the fact that this population does not represent the general population, it also comprises a demographic group of ill-formed "households." Most students are only semi-independent households with mixed income sources and unclear and evolving budgetary responsibilities. However, the sample also has advantages: the sample is relatively large with 180 observations, and the sample population is relatively representative of a target audience of young adults attending college – an interesting group for the topic of interest. The second concern with the study is inherent to the nature of the topic. The likelihood and effects of global warming are uncertain and laden with long time lags. One must use some caution in judging whether any population can formulate meaningful preferences over polices concerning global warming.

We have chosen to view these study features as creating an opportunity. If we can find evidence of meaningful willingness to pay signals in these circumstances, we think that the case for using contingent valuation studies to estimate policy values is bolstered. Moreover, the circumstances of the case produce some edifying results that should generate fruitful topics for future studies.

The study produced three principle results. First, despite the respondent's characteristics and the difficulties of the topic, we were able to find a clear signal of willingness to pay. Results showed that the coefficient on the willingness to pay variable was clearly significant and appropriately signed for all policies: increases in the price reduced the probability that respondents would approve a policy. Considering that we could not find a meaningful income/budget value and that the marginal effect of price was found to be very low, we think it quite possible that the value levels obtained in the study are incorrect – presumably high. However, the price variable was clearly not an arbitrary feature to the respondents. Despite a considerable heterogeneity of preferences over "green" policies, respondents seemed to respond to the price variable in a consistent and commensurable manner. That is, for our overall, sample money does matter – even for those backing the most "green" policy.

We also found that, after controlling for prices, preferences over the emissions-reduction policies was very heterogeneous. Respondents did not rank the policies simply on the basis of their degree of effectiveness in reducing carbon emissions. This leads to our two other major conclusions. First, we find evidence that "green attitudes" affect preferences among the policies. Other things equal, some respondents favor one policy over another. There appears to be an underlying "base value" focused on a "moderate" policy of 25% emissions reduction as indicated by the coefficients on the policy alternatives. From this base, the more green one's attitudes are, the more likely one is to favor stronger policies and vice versa.

We also find evidence that respondents imputed attributes to the policies in addition to the emission reduction attribute explicitly presented in the study. We hypothesize that there may be at least two additional policy dimensions: predicted policy effectiveness and predicted "life-style" costs of the policy. By life-style costs we mean changes in standards and manner of living that are not completely captured by the monetary cost/price assigned in the study. In our expanded model we found that those most willing to drive less to help reduce global warming were most likely to approve a stronger carbon reduction policy. We think finding evidence of a lifestyle cost dimension is a very important result that may apply to many contingent valuation studies in a way that is analogous to how time affects travel cost estimates. In travel cost studies the monetary expenditures do not capture the entire opportunity costs of travel. Trips with the same monetary expenditures may require different times. Similarly, environmental policies may have non-expenditure costs that are not well captured by the prices assigned by the researcher. We must remember that contingent valuation estimates are truly contingent on the hypothetical policy package constructed. Respondents will respond to all dimensions of the policy package -- including perceived policy provisions that may alter their expected way of life.

References

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

Boyle, K.J., H.F. Maconald, H. Cheng, and D.W. McCollum. "Bid Design and Yea Saying in Single-Bounded, Dichotomous Choice Questions." Land Economics 74(February 1998): 49-64.

Brown, Champ, Bishop, and McCollum. "Which response format reveals the truth about donations to a public good." *Land Econ.* 72(May 96): 152-166.

Cline, William R. *The Economics of Global Warming.* Institute for International Economics: Washington, D.C. 1992.

Cooper, J. and J. Loomis. "Sensitivity of Willingness to Pay Estimates to Bid Design in Dichotomus Choice Contingent Valuation Models." Land Economics, 68(May 1992): 211-24.

Diamond, P.A., and J.A. Hausman. "On Contingent Valuation Measurement of Nonuse Values," in J.A. Hausman (ed.) Contingent Valuation: A Critical Assessment. NY: North-Holland, 1993.

Dillman, Don. *Mail And Telephone Surveys: The Total Design Method.* Wiley: New York. 1978.

Dunlap, Riley. "The Quality of Washington's Natural Resources: A 1990 Survey of Citizen Opinions." Social and Economic Sciences Research Center: Washington State University. 1990 (unpublished).

Dunlap, Riley, G.H. Gallup, Jr. and A.M. Gallup. *The Health of the Planet Survey: a preliminary report on attitudes on the environment and economic growth measured by surveys of citizens in 22 nations to date.* George H. Gallup International Institute: Princeton, New Jersey. May 29, 1992.

Dunlap, Riley and Rik Scarce. "Trends: Environmental Problems and Protection." *Public Opinion Quarterly* 55 no. 4, Winter 1991. 651-672.

Econometric Software, Inc. "LIMDEP Version 7.0" <http://www.limdep.com> Econometric Software, Inc. (accessed: December 4, 2000)

Ekins, Paul. "Rethinking the Costs Related to global Warming: A Survey of the Issues." *Environmental and Resource Economics*, 6. 1995. 231-277.

Epley, Nicholas and David Dunning. "Feeling 'holier than thou': are selfserving assessments produced by errors in self or social prediction?" *Journal of Personality and Social Psychology.* 79(Dec 2000): 861-875.

Fletcher, Susan R. "Global Climate Change Treaty: Summary of the Kyoto Protocol" <http://www.cnie.org/nle/clim-3.html> Congressional Research Service: Report for Congress. Committee for the National Institute for the Environment. March 11, 1998. (Accessed: June 16, 1998)

Greene, William H. *Econometric Analysis: Third Edition.* Prentice-Hall, Inc: Upper Saddle River, New Jersey. 1997.

Intergovernmental Panel on Climate Change (IPCC). "Climate Change 2001: The Scientific Basis," Third Assessment Report; IPCC Working Group III, United Nations Environment Programme. Summaries for policymakers located at: <http://www.usgcrp.gov/ipcc/wg1spm.pdf> 2001.

Kahneman, D. and J.L. Knetsch, "Valuing Public Goods: the Purchase of Moral Satisfaction," JEEM, 21(1992): 57-70.

Lareau, Thomas J and Douglas A. Rae. "Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique." *Southern Economic Journal,* 55 no. 3 (January 1989): 728-742.

McFadden, Daniel. "On Conditional Logit Model of Qualitative Choice Behavior." In: Frontiers of Econometrics. Paul Zarembka, Editor. Academic Press: New York. 1974.

Mitchell, R.C. and R.T. Carson. 1989. *Using Surveys to Value Public Goods.* Washington DC: Resources for the Future.

[i] This means that our elicitation formats were not very efficient, which is also indicated by the very rare selection of the status quo choice. In principal, one should try to get a distribution of responses centered on the eventually estimated WTP (see, e.g. Cooper and Loomis, 1992). In the current study we used actual estimated costs according to Cline for our bid values because we thought this had interesting policy implications.

Appendix Table: Effects of emission reductions on global warming impacts

Global CO2 Reduction levels	Global Temperature Increase	Potential Risks and Impacts
60%	0 Celsius	Stops further atmospheric build-up of greenhouse gases. Maintains climate much as it is today.
50%	0.2 Celsius (0.4 F)	Few changes in climate. Minor increases in desertification. Lowers agricultural production in some parts of the world.
25%	1 Celsius (1.8 F)	Global rise in sea level up to 6 inches. Flooding of low lying coastal regions. Reduced agricultural production in many parts of the world. Increases in diseases such as Malaria.
7%	2.2 Celsius (4 F)	Sea level rise of about 1 foot. Flooding of coastal regions. Changes in agricultural production. Increased frequency and intensity of storms, such as tornadoes and hurricanes.
No reductions: Business as Usual	3 Celsius (5.4 F)	Over 1 and a half feet rise in sea level. Coastal regions and low lying islands flooded. Loss of fresh-water reserves in glaciers and ice caps. Large increases in the incidence of floods and droughts. Northward shifting of agricultural patterns. Loss of high altitude ecosystems. Increased ecological instability. Expected increases in political instability in poor countries due to food and health problems.

What is the Value of a Bird? A Proposed Methodology Based on Restoration Costs and Scarcity by Species

Steve Hampton, Ph.D.
Resource Economist
Department of Fish and Game
shampton@ospr.dfg.ca.gov
(916) 323-4724

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Abstract: This paper begins by examining the recent history of natural resource damage assessment (NRDA), focusing on the shortcomings of recent methodologies and the need to develop simplified and reasonable methods for estimating the value of individual birds. A theoretical model is proposed, where the value of a bird is a function of known restoration costs for certain species, the relative scarcity of each species, the current population trend of the species, and other factors. Finally, an preliminary (and incomplete) example is presented, using the model to estimate restoration costs per bird for regularly occurring species in California.

Background of NRDA and Bird Kills

In the aftermath of an oil spill or other pollution event, various federal and state statutes, such as the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA or Superfund), the Oil Pollution Act of 1990 (OPA), and California=s Lempert-Keene-Seastrand Oil Spill Prevention and Response Act (Government Code " 8670 et seq.), authorize trustee agencies to seek monetary compensation for injured natural resources. The Department of the Interior, responsible for promulgating NRDA regulations pursuant to CERCLA, has suggested that Acompensable value@ due to the public should Aencompass all of the public economic values associated with an injured resource, including use values and passive-use values such as option, existence, and bequest values@ (56 Federal Register 19760 (1991)). California Fish and Game Code ' 2014 specifies that Athe state may recover damages in a civil action against any person or local agency which unlawfully or negligently takes or destroys any bird, mammal, fish, reptile, or amphibian...@ and that Athe measure of damages is the amount which will compensate for all the detriment...@

Since the Exxon Valdez oil spill in 1989, NRDA has evolved into a well-defined discipline, with its own case history, legal precedents, and economic literature. While debates over contingent valuation were waged in many quarters, the National Oceanic and Atmospheric Administration (NOAA), responsible for promulgating NRDA regulations pursuant to OPA, released federal guidelines in 1996, recommending that the measure of damages be based on restoration costs. That is, the public may be compensated for the interim lost use of the resources based on the cost to supply (or re-create) the amount of natural resource services that were lost due to a pollution event. This Acompensatory restoration@ is not to be confused with Aprimary restoration@, which are post-cleanup actions at the site of the incident designed to speed the recovery of the impacted resources. NOAA recommends Habitat Equivalency Analysis (HEA) as a preferred method for calculating damages (see Mazotta et al. 1994, Unsworth and Bishop 1994, and NOAA 1995 for details on this method).

This recommendation quickly shifted the basis for valuing natural resources from the demand side to the supply side. The question was no longer Ahow much does the public value this resource?@ but Awhat does it cost to create, enhance, or restore this resource?@ It seemed as if economists using such valuation methods as contingent valuation and travel cost analysis presumably could be replaced by restoration ecologists and accountants using the annuity formula, at least in cases that did not involve active human uses.

Aside from the philosophical questions raised by this methodological shift, a host of practical questions have arisen. HEA is well-suited for injuries to habitats for which there are known restoration alternatives and well-documented costs. However, there are many pollution events that have negligible impacts to habitat but kill a large number of animals. The most common example of this is an oil spill into open water where the oil moves into the water column and atmosphere, and/or is cleaned up. Thus, it impacts no shoreline habitat but often kills many birds that contacted the oil while it was on the water=s surface.

Those conducting the NRDA of the North Cape oil spill faced this problem, but were largely limited to two species of birds (loons and eiders). They successfully adapted HEA into a bird resource equivalency analysis (REA) and calculated the number of lost loon and eider years due to the spill (Sperduto et al. 1999). They further benefitted from the existence of restoration data, so that they could estimate the potential gain (in Abird years@) from a proposed restoration project. The projects could then be scaled to provide the appropriate amount of compensation.

However, in many spills a wide variety of species of birds are impacted, such that often no one species accounts for even the majority of the birds killed. In this situation, replicating the North Cape bird REA for each and every impacted species would be both time consuming and expensive, and undoubtedly limited by a lack of data regarding potential restoration benefits. The dilemma is the fact that we have restoration project data (regarding both the benefits and costs) for only a handful of the 650 or more bird species that regularly occur in the United States. It is thus impractical to develop a REA for all of these species. Additionally, responsible parties desire to minimize assessment costs and avoid lengthy and expensive studies.

The Type A model, developed to model the impacts of oil spills and to calculate the damages, anticipated this situation, where values per bird would be required. In cases where habitat restoration fails to address impacts to certain species, the Type A model flips from a supply-based valuation method to a demand-based method, relying on a Aprice list@ for birds by species. These values were estimated based on the likelihood of viewing a species from shore, and vary from region to region. However, the resulting list of values per species is at best unsatisfying. In many cases, the results are baffling to both the restoration ecologist and the avid bird watcher. A Tundra Swan, rather common inland in appropriate habitats, is listed at over $13,000 per bird year in one region, while the Common Murre, a seabird struggling to recover from decades of human-induced population impacts, is valued at just over $1 per bird year.

A Proposed Model

This paper focuses on this problem, addressing the practical question of how to estimate restoration costs for all species of birds. The basic idea is to identify well-documented restoration benefits and costs for a few species and to extrapolate those values to all species based on the relative scarcity of the species and other factors.

An underlying question is the correlation between the value of a bird, as measured by public demand, and the costs to Asupply@ a bird, as measured by restoration costs. Certainly the NOAA recommendations presume this correlation. The key premise to this model is that public value and restoration costs are correlated with each other, and that both are negatively correlated with the population status of the species. That is, with some caveats, scarce birds are more valued by the public and are more expensive to restore, and common species are less valued and are cheaper to restore.

With respect to the supply side, is it truly cheaper to restore common birds, and more expensive to restore rare birds? What little data there is does suggest this correlation, as will be presented in the forthcoming example. Rare species typically have narrow and specific habitat preferences, often in conflict with human activities. The Spotted Owl=s affinity for old growth coniferous forests is an obvious example. With this species, a typical restoration alternative is the acquisition of land. The opportunity cost of preserving such habitat, which also has a high commercial value, is thus quite high. At the other end of the spectrum, common species are often adapted to a wide variety of habitats, including those with considerable human disturbance. Anyone seeking to restore one of these species may have a wide variety of potential restoration options, some of which may be relatively inexpensive (and involve little opportunity cost). However, there will certainly be individual species, based on their ecological preferences, that do not fit this overall pattern.

Note that bird restoration projects may be quite variable in their design and goals. They may seek to increase fledgling success rates or even simply expand the number of nesting birds.

Such goals may be achieved via efforts to reduce human disturbance or animal predation of nests, or by somehow improving the nesting habitat by altering the vegetation or landscape in some way. Other projects may seek to increase juvenile or adult survival rates on the breeding or wintering grounds by implementing measures to minimize whatever is killing the birds. These are only some examples.

With respect to the demand side, we can examine the behavior of active bird viewers (also known as Abird watchers@ or Abirders@). Such people may vary in their willingness to pay to view birds. We can think of the casual backyard birder with a bird feeder, the occasional birder who visits a wildlife refuge or attends a birding festival a few times a year, or the hardcore birder who flies across the country on a moment=s notice to view a rarity. The ends of the spectrum illustrate the basic principle: that the value of a bird is a function of the frequency of it being seen. Small expenditures are paid to see common birds; large expenditures for rare species. However, using scarcity as the basis for the value of a bird ignores other attributes that may also contribute to a bird=s value. Larger and/or more colorful species have more charisma, and may have greater value for some people. Game birds (e.g., ducks, geese, pheasants) have a consumptive value, as well as non-consumptive value. Alternatively, a few species (e.g., crows, gulls) may be viewed negatively by some.

Even if we accept the notion that scarce resources are more valuable, when it comes to natural resource management, scarcity may not communicate the entire status of the population. Some bird species (e.g., Heermann=s Gull, Elegant Tern, and Brown Pelican) are rather common, yet highly vulnerable because they nest at only a few locations. Perhaps a Avulnerability index@ that incorporates scarcity as well as other parameters (such as number of breeding colonies and population trends) would provide a more complete measure of population status.

One additional problem is the definition of scarcity. A bird may be scarce in one region, yet quite common elsewhere. Thus, the geographic area of reference becomes relevant. For example, an oil spill off California may kill a Long-tailed Duck (formerly known as Oldsquaw), which is rather rare there but much more common in other parts of North America. Since species are listed as threatened or endangered, as well as managed by trustee agencies, at a more regional or statewide level (in order to protect regional sub-populations), I would suggest a state or regional level as the maximum geographic unit when defining scarcity, but with care exercised in applying the results to species like the Long-tailed Duck in California. At the other end of the spectrum, there is often bird abundance data at the county level as well, which may be used in defining scarcity. This geographic unit may be so small, however, that a pollution event may transcend its boundaries. Certainly, the non-consumptive users (bird watchers) that contribute to the bird=s value can travel from outside a county to watch wildlife and do so quite regularly. Additionally, a restoration project seeking to benefit birds may not be entirely within the county where the pollution event occurred.

All these caveats aside, we will forge ahead, using the premise that value, and thus restoration costs, are correlated with scarcity. More formally, we propose that restoration costs (C) are a function of the overall size of the population (Q). Thus, we have

$$C(Q) \text{ where } dC/dQ < 0.$$

Further more, let us hypothesize a relationship between cost and population size, whereby costs increase exponentially as a species is more scarce ($d^2C/dQ^2 > 0$). Figure 1 illustrates this simple principle and the resulting curve is somewhat intuitive. The most common birds are relatively inexpensive to restore, while the rare species may involve exponentially increasing expenses. One can think of the millions of dollars spent to help the California Condor recover from the brink of extinction.

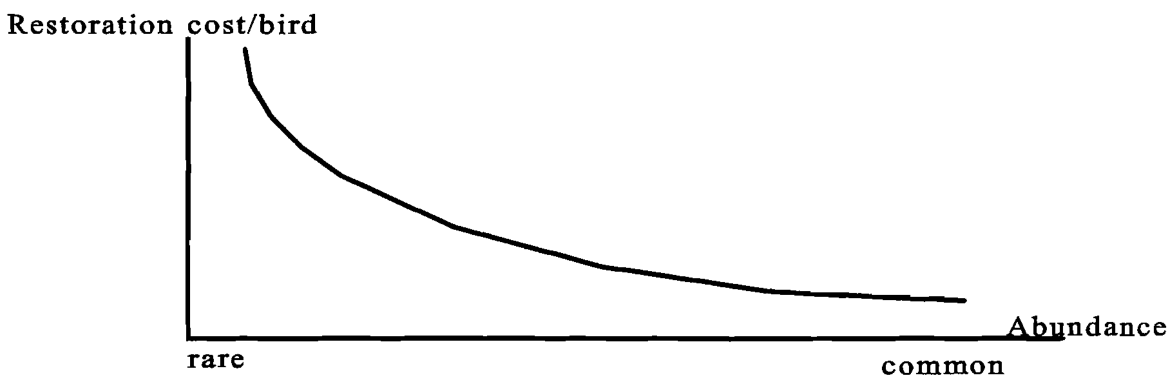


Figure 1: Restoration Costs for Each Species

Note that the costs per bird from these restoration projects are not simply based on the cost of the project divided by the potential number of new birds or new bird years created. Rather, they are derived from REAs of the lost bird years due to a bird kill and the bird years gained from the restoration project. The cost of the REA-scaled project is then divided by the original bird kill, giving us the true cost per bird killed in the incident.

This approach will create a discrepancy between species with steady (or increasing) populations and those with declining populations. REA calculates lost bird years, factoring in the natural recovery time of the species and thus calculating the interim loss. For a species whose population is steady (and assuming the incident does not cause a catastrophic population-level impact), we might assume that only one or two generations are lost, or even that only one year=s worth of birds are lost (see Sperduto et al. 1999). (How many generations and/or years to carry out injury in a REA is a subject of another paper in progress.) Beyond one or two generations, we assume the population will naturally recover from the incident and no additional losses are incurred. For a species with a declining population, lost birds will never be replenished naturally and are therefore lost into perpetuity. In this case, the number of lost bird years per individual killed is disproportionately higher than for other species. In fact, some preliminary REA results for certain species suggest the number of lost bird years per individual killed may be five to ten times greater. Thus, our model, as it stands now, estimates the REA-based restoration costs of a group of birds treated in a uniform manner with respect to the length of impact (e.g., one generation vs. into perpetuity).

If we simply divide all species into those that would suffer a one generation impact and those that are lost into perpetuity, an appropriate and simple addition to our model would be a dummy variable (D) for species with declining populations. For these species, the REA assumes a bird is lost into perpetuity. The equation now becomes

$$C(Q,D)$$

where $dC/dD > 0$. That is, REA-based restoration costs per injured bird increase if the species is declining.

Formally, let us take the most simple route (mathematically) and assume that the elasticity (b) of cost with respect to abundance is constant. For every percentage change in the abundance of a species, there is a fixed percentage increase in its restoration costs. We can employ a simple demand-style function, where:

$$Q = a \cdot C^b$$

where a is a constant and b is the elasticity between cost and population size. D, the dummy variable for declining species, simply shifts our curve up on Figure 1. In order to estimate C for each species, all that is required is Q for each species and enough data points of C to enable us to estimate the values of a and b.

Note that the proposed functional form will produce a substantial confidence interval for the rarer species, simply because the curve is shallow for the common birds but quite steep for the rarer species. Thus, for the rare species the predicted restoration costs will be quite sensitive to the parameters a and b. For example, if the curve was fit so that a very common bird has a restoration cost of $120/bird and we assume an elasticity (b) of -1.0, a = 114. If we assume that b = -0.5, a = 10.4. In the first case, a moderately common bird may have a restoration cost somewhere between $200 and $300 per bird, depending on which of the two sets of parameters we choose. However, for a rare bird, the estimated restoration costs will range from $500 to $2,500 per bird. This problem may be irrelevant, however, as case-specific REAs are likely to be developed for any NRDA involving rare species. It is primarily for the more common species that we need a tool for estimating restoration costs.

Preliminary Results from California

We have begun applying the proposed model to California. The first step was to acquire a measure of the abundance of each of the regularly occurring species in the state. For the purposes of this exercise, we have developed an Aabundance index@ using DeSante and Pyle (1986) and county bird checklists. The details of this index, as well as suggestions for improvements, are explained in Appendix A (not available in this Working Draft; contact the author for details). For each species, we have created an abundance index (Q in our model), ranging from 0 to 1, from rare to common respectively.

The second step is to identify REA-based restoration costs per bird for as many species as possible. Ideally, the data will include a mix of common and rare species. Care must be taken as many restoration projects target entire habitats or areas, not just a single species. Nevertheless, there are well documented projects in California for Mallard nesting habitat and for Common Murre nesting habitat. Estimating the REA-based costs for California Condor restoration would be an endeavor unto itself, but suffice it to say the number is possibly several hundred thousand dollars per bird. Simply looking at these three examples lends support to our hypothesized exponential function, as depicted in the curve in Figure 1.

Table 1: Species with Known Restoration Costs (based on REA Calculations)

SPECIES	ABUNDANCE INDEX	RESTORATION COST/LOST BIRD
Mallard	.9500	$120
Common Murre	.2948	$600
California Condor	.0001	$?00,000

[Note: research into restoration costs for other species is currently on-going.]

Note that this exercise may be extended to rarely occurring species in the state, such as vagrants from the eastern U.S. or from Asia. The results, however, are largely erroneous, as suggested in the previous discussion of scarcity with respect to geographic unit. The abundance index values for these species are so low that the corresponding estimated costs per bird are in the millions of dollars. The true costs of restoring these species, especially in the regions where they normally occur, is undoubtedly much lower. Thus, lines must be drawn, not just around the geographic unit but around the appropriate species to consider in the analysis.

Another preliminary finding is that Western Bluebirds are unexpectedly cheap to restore. With an abundance index of .7000 in California, a curve fit to the Mallard and Common Murre above would estimate bluebird restoration costs at approximately $180 per bird. However, a specific REA-based analysis of a restoration project for bluebirds (involving the creation and monitoring of bluebird nest boxes over a 12-year period) estimated the costs at $30 per injured bird. It may be that songbirds and similar species are cheaper to restore than larger birds such as waterfowl, seabirds, and raptors, who generally nest in wetlands or other sensitive habitats. These larger birds are generally K-adapted species with slow reproductive capacity and high annual survival rates (except waterfowl), whereas songbirds are closer to r-adapted species that produce many juveniles per year but have low annual survival rates. Further research will explore this issue.

Conclusions

As methods for NRDA increasingly use restoration costs, rather than consumer valuation, as their basis, knowledge of potential restoration costs are at a premium. However, scarce data and a lack of restoration history for many species limit the applicability of species-specific REAs to scale compensatory restoration projects. This proposed method attempts to utilize the limited data that exists regarding bird restoration and to extrapolate that to other species. The goal is thus to estimate the potential restoration costs of these species, or at least to approximate the supply-side driven value that may be associated with them. Given the inherent variability in wildlife restoration, more data is needed to complete the application of this model to California birds, but preliminary information suggests it may be worthwhile.

This model may then be used in expedited NRDAs to minimize assessment costs and estimate restoration costs, especially in cases where a wide variety of species have been impacted.

References

DeSante, D. and P. Pyle. 1986. *Distributional Checklist of North American Birds. Volume I: United States and Canada.* Artemesia Press: Lee Vining, CA.

Mazotta, M.J., J.J. Opaluch, and T.A. Grigalunas. 1994. Natural resource damage assessment: The role of resource restoration. *Natural Resource Journal* 34: 153-178.

NOAA. 1995. *Habitat Equivalency Analysis: An Overview.* NOAA Damage Assessment and Restoration Program. Department of Commerce.

Sperduto, M., C. Hebert, M. Donlan, S. Thompson. 1999. *Injury Quantification and Restoration Scaling for Marine Birds Killed as a Result of the North Cape Oil Spill.* United States Fish and Wildlife Service.

Unsworth, R.E. and R.C. Bishop. 1994. Assessing natural resource damages using environmental annuities. *Ecological Economics* 11: 35-41.

The Public's Willingness to Pay for Improving California's Water Quality

Douglas M. Larson

Daniel K. Lew

and

Yuko Onozaka
Department of Agricultural and Resource Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Abstract: This paper discusses the preliminary results of a CVM mail survey of California households. A sample of California households were asked whether they would support a water quality program that would improve the water quality in lakes, rivers, streams, coastal waters, wetlands, and estuaries throughout the state to levels at which basic uses are supported. A referendum with follow-up question format was used and a linear willingness-to-pay (WTP) equation was estimated. WTP is modeled as a function of demographic and attitudinal variables. Preliminary results suggest California households would be willing to pay approximately $2 billion per year to fund efforts to improve the state's water quality to levels that would be in compliance with current water policies.

Introduction

California has been, and continues to be, a magnet for population and industrial growth. The problems this poses for environmental quality have been well-documented, particularly for air pollution. Less well-publicized, perhaps, are the problems California experiences with water quality. While many sources of water pollution have been identified and are being successfully regulated under the state's Porter Cologne Act and the federal Clean Water Act (PL 92-500), water quality problems persist and are severe in some localized areas, primarily as a result of non-point source pollution. In fact, the most recent assessment of the state's water quality revealed that a large percentage of water bodies throughout the state are either impaired or in danger of becoming impaired (SWRCB, 1999).[1]

The leading sources contributing to the degradation of California's water quality are agricultural runoff, forestry activities, storm water runoff, storm sewers, and unspecified non-point sources (U.S. EPA, 2000). An estimate of the money required to control just one of these sources, storm water runoff, throughout the state puts the cost in the tens of billions of dollars per year (Stanley Hoffman and Associates, 1998). Given the growing list of other local, state, and national concerns competing for limited public funds, it is useful to explore the public's interest in this issue. More specifically, in deciding how far California should go to improve its water quality, it is necessary to know how much the people of California value water quality improvements. At present, there is relatively little information available on this important topic at the state level. However, it does become important as private and public entities within the state struggle to meet the requirements of state and federal clean water laws.

This paper presents preliminary results from a contingent valuation survey of California households intended to gauge the public's willingness to pay (WTP) for improving water quality in California to a level commensurate with the goals of the Clean Water Act. A common approach for extrapolating sample results to the population in CVM studies is to estimate WTP as a function of characteristics that may differ between the sample and population (Harrison and Lesley, 1996; Loomis, 1987; Rubin, Helfand, and Loomis, 1991; Mitchell and Carson, 1989). Economic theory suggests variables that are likely to be important determinants of willingness to pay, among them income, prices, and taste variables. Demographics are a common set of characteristics used as taste indicators (Loomis, 1987). One of the goals of this study is to explore the use of an attitudinal variable as a covariate in the WTP specification that can be used to adjust the WTP for the sample to the population. This is made possible by the fact that a measure of this variable for the population is available from an independent general population survey. To our knowledge, previous applications have not attempted to adjust WTP using attitudinal variables for which there is a population counterpart obtained through a survey of the population of interest.

Other efforts to value the public's WTP for water quality in the existing literature include Carson and Mitchell (1993), Bockstael, McConnell, and Strand (1989), Oster (1977), and Gramlich (1977) to name a few. The approach taken here departs from previous efforts in both scope and depth. Whereas Bockstael, McConnell, and Strand, Oster, and Gramlich measured the public's WTP for swimmable water quality in specific water bodies (the Chesapeake Bay, Merrimack River Basin, and Charles River, respectively), our goal is to measure the WTP for improving water quality of an agglomerate of water bodies as Carson and Mitchell did. But unlike Carson and Mitchell, we focus on California instead of the entire nation and estimate the value of a single water quality improvement, from the status quo to the level where all the designated uses for each water body are unimpaired. This differs from the treatment by Carson and Mitchell (and Mitchell and Carson [1989]), who estimate the national WTP to raise the minimum water quality in almost *all* (99% or more) freshwater bodies from their present levels to boatable, fishable, and swimmable levels regardless of the designated uses for these waters.

The Survey

To begin understanding how Californians feel about water quality in the state, a pilot contingent valuation survey was developed. The primary purpose was to collect information from Californians on their valuation of water quality in the state. The final survey was divided into three sections. The first part asked for information on the individual's attitudes towards water quality and the environment; the second consists of an explanation of the good being valued and a referendum-type valuation question; and the final section consists of demographic questions.

In the first section, respondents were asked for their opinions about water quality issues, attitudes toward clean water, reasons why they may or may not value water quality, and whether or not they participated in recreation activities at or near different types of water bodies.[2] The section included two questions eliciting concern for environmental issues that are identical to the questions asked by the Field Institute in a California Field Poll (Field Institute, 1997).[3] The two questions asked in the 1997 California Field Poll were the following:

- "How concerned are you about air and water pollution?" and
- "How concerned are you about protecting the environment?"

Possible responses were "extremely concerned," "somewhat concerned," "not too concerned," "not at all concerned," or "not sure." The exact wording used in the Field Poll was preserved in the survey. Viewed as indicators of an individual's preference (taste) for environmental goods, the use of responses to these questions can be used to explain the variation of WTP and thus be used in the construction of population WTP estimates.

The second section presented information on the public good being valued, improvement of water quality in the lakes, rivers, streams, estuaries, wetlands, and coastal waters throughout the state to levels that support each water body's designated beneficial uses. After respondents were informed of the current status of water in California, a proposed program was described that would raise water quality in water bodies throughout the state to levels that would be in compliance with state and federal clean water laws (levels that would protect beneficial uses). Respondents were then told the total cost of the proposed program, and that it would be implemented should a majority of voters express support for the program.

Following the program explanation, respondents were told that the program would have to be paid through an additional surcharge on households' water utility bills. Renters who do not pay for their water bill would see an increase in their rent.

The contingent valuation (CV) question is structured in a referendum with follow-up (also called the double-bounded dichotomous choice) format, where individuals are asked to accept or reject an initial amount and, depending on this first response, asked to accept or reject a higher (if they accepted the initial amount) or lower amount (if they rejected the initial amount). Compared to a format that only asks a single valuation question, this double-bounded format has been shown to increase the efficiency of estimated WTP parameters (Hanemann, Loomis, and Kanninen, 1991).

Respondents were asked whether they would support the program if the monthly surcharge amount was some amount $\$B_I$ and reminded that paying the surcharge would decrease the amount of disposable income available for other purchases. Those who answered "yes" to $\$B_I$ (they would pay the amount) were asked whether they would be willing to support the program if it cost a higher amount, $\$B_H$, while those answering "no" to $\$B_I$ (they would not pay the amount) were asked if they would support the program if it cost a lower amount, $\$B_L$. Respondents who said "no" to both amounts were then asked if they would be willing to pay $1. Individuals answering "no" to all questions were asked why their response had been "no" to all the preceding surcharge amounts to isolate protest zeros from those individuals who legitimately did not have a positive willingness-to-pay for the program.

The final section contained demographic questions that could potentially be used to help explain the variation in observed WTP. These variables included household income, age, educational attainment, gender, household size, and employment status. In addition, individuals were asked how long they resided in California and whether they had ever or are currently members of an environmental or conservation organization. For the purpose of this preliminary analysis, household income and education were used as WTP shifters.

A Double-Bounded Dichotomous Choice Model

Because improvement in water quality is an economic good, we expect willingness to pay to be non-negative. Thus we model individual i's willingness to pay as a latent variable, WTP_i^*, which can be expressed as

(1) $$WTP_i^* = x_i \cdot \beta + \sigma \cdot \varepsilon_i .$$

(2)

The observed willingness to pay, WTP is

$$\begin{aligned} WTP_i &= WTP_i^* && \text{for } \varepsilon_i \geq (WTP_i - x_i \cdot \beta)/\sigma \\ &= 0 && \text{otherwise} \end{aligned}$$

where $\boldsymbol{x}_i$ is a vector of explanatory variables that are observable, ε_i is a symmetric disturbance term with mean 0 and unit variance, and $\boldsymbol{\beta}$ and σ are parameters to be estimated.

We would expect to observe individual i supporting the program (agreeing to pay surcharge) if the surcharge amount was less than or equal to their WTP. Conversely, if the surcharge is greater than the individual's WTP, then we would observe the individual rejecting the proffered amount. In terms of probability statements, this can be formalized as follows:

(2) $$\Pr(Yes) = \Pr(WTP_i \geq B_{i1}),$$

where B_{i1} is the initial monthly surcharge (or "bid") confronting the individual (amount of surcharge in the first referendum question).

If we were to use only the responses to the first bid B_{i1} in a single bounded model, the probability of observing support for the program by individual i could be written as

(3) $$\begin{aligned}\Pr(Yes) &= \Pr(WTP_i \geq B_{i1}) \\ &= \Pr(x_i \cdot \beta + \sigma \cdot \varepsilon_i \geq B_{i1}) \\ &= 1 - \Pr\left(\varepsilon_i \leq \frac{B_{i1} - x_i \cdot \beta}{\sigma}\right) \\ &= 1 - F\left(\frac{B_{i1} - x_i \cdot \beta}{\sigma}\right)\end{aligned}$$

where $F(\cdot)$ is the cumulative distribution function of ε_i. If ε_i follows a standard normal distribution, then $F(\cdot)=\Phi(\cdot)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution, and (3) can be estimated with standard probit techniques.[4]

Of course, using the additional information contained in the responses to the follow-up surcharge amounts result in gains in efficiency of the estimated WTP parameters (Hanemann, Loomis, and Kanninen). However, the introduction of a follow-up question to the survey may introduce response effects. Different WTP amounts estimated from the same data but employing alternatively a single-bounded model and a double-bounded model suggest the potential for question format biases (Herriges and Shogren, 1996).

Several hypotheses to explain this phenomenon have been put forth, including response incentive effects and anchoring effects.[5] Cameron and Quiggin (1994) suggest the possibility of different underlying valuation functions associated with the first and second valuation questions resulting from the creation of incentives to respond to the follow-up question in a manner dependent upon the initial referendum question. For example, they suggest that respondents who reject the first offered amount may feel guilty and thus may feel obligated to respond more positively to the second offered amount. Another possibility is that respondents may answer negatively to the second amount because they may become frustrated because they feel like the researchers are attempting to either "eke at least some money out of them by lowering the bid" (Cameron and Quiggin, 1994, 228) or, conversely, drain them of more money just because they answered favorably to the first amount.

Alberini, Kanninen, and Carson (1997) hypothesize other response incentive effects consistent with multiple underlying valuation functions. They suggest that in light of the additional cost information provided in the follow-up question, respondents perceive a different policy package or scenario than was originally conveyed to them. Because the initial cost of the program implies the program can be delivered at that amount, individuals confronted with a higher follow-up cost (those who said "yes" to the initial amount) may view this as a sign of government waste and consequently answer "no" to the higher amount. By the same logic, individuals who had rejected the initial amount may view the lower amount as associated with a scaled-down version of the program described to them. To account for these factors, they suggest modeling WTP separately for each valuation question. However, in practice, a difficulty arises in determining the appropriate valuation function for policy decisions, the valuation function associated with the initial referendum question or the one for the follow-up question. Unfortunately, economic theory and the CV literature do not provide a clear-cut manner to decide between the two.

An alternative explanation for why incorporation of follow-up responses in estimation may lead to different WTP amounts is offered by Herriges and Shogren (1996). They assert that responses to follow-up questions are subject to starting point, or anchoring, bias (Mitchell and Carson, 1989). According to this view, the initial amount offered individuals influences their responses to subsequent valuation questions by acting as an anchor from which uncertain individuals can base their responses. Therefore, what the researcher observes in follow-up valuation questions is a mixture of the individual's true underlying preferences (WTP) and the effect of the initial amount offered to them in the previous question. McLeod and Bergland (1999) offer a similar view, but motivate the effect of the initial bid in a Bayesian framework and view the problem as one of updating the model to include priors derived from the individual's experience with the first valuation question.

For this paper, the effects of the initial bid are modeled similarly to Herriges and Shogren. To account for potential starting point bias, we model the "observed" WTP, as

$$WTP_i^{obs} = \left(\frac{1}{\gamma + 1}\right) \cdot WTP_i + \left(\frac{\gamma}{\gamma + 1}\right) \cdot B_{i1}, \tag{4}$$

where we would expect that $\gamma \geq 0$, in which case $\gamma/(\gamma+1) \in [0,1)$ and measures the extent of anchoring, or the proportion of the observed WTP response that is attributable to the bid the individual was offered. If $\gamma/(\gamma+1)$ is 0, there is no starting point bias and WTP^{obs} = WTP, the true underlying preference function defined by (1). Conversely, as $\gamma/(\gamma+1)$ approaches 1, the effect of the initial bid dominates the true WTP in the observed WTP response. Herriges and Shogren (1996) show that a positive $\gamma/(\gamma+1)$ unequivocally widens the individual's probability regions, thus reducing the gains to efficiency resulting from the added information provided by the follow-up responses.

Responses to the two valuation questions in the referendum with follow-up question format fall into four categories: (1) those who reject, or say "no," to both surcharge amounts (NN)[6]; (2) those rejecting the initial amount and accepting the follow-up (YN); (3) those saying "yes" to the first amount and "no" to the second (NY); and (4) those saying "yes" to both amounts. Assuming the disturbance terms follow a standard normal distribution, and using the linear specification in equation (1) in equation (4), the probabilities of observing each response pattern for individual *i* are

$$\Pr(NN_i) = \Phi\left(\frac{(\gamma + 1) \cdot B_{iL} - \gamma \cdot B_{i1} - x_i \cdot \beta}{\sigma}\right), \tag{5}$$

(6) $$\Pr(NY_i) = \Phi\left(\frac{B_{i1} - x_i \cdot \beta}{\sigma}\right) - \Phi\left(\frac{(\gamma + 1) \cdot B_{iL} - \gamma \cdot B_{i1} - x_i \cdot \beta}{\sigma}\right),$$

(7) $$\Pr(YN_i) = \Phi\left(\frac{(\gamma + 1) \cdot B_{iH} - \gamma \cdot B_{i1} - x_i \cdot \beta}{\sigma}\right) - \Phi\left(\frac{B_{i1} - x_i \cdot \beta}{\sigma}\right), \quad \text{and}$$

(8) $$\Pr(YY_i) = 1 - \Phi\left(\frac{(\gamma + 1) \cdot B_{iH} - \gamma \cdot B_{i1} - x_i \cdot \beta}{\sigma}\right),$$

where B_{iL} is the lower amount offered to individual i if she rejects the initial amount, B_{iI} is the initial surcharge amount offered i, and B_{iH} is the higher amount offered to i should she say "yes" to the initial amount. Equations (5)-(8) define the four regions of the CDF of the error distribution for individual i. The log-likelihood function is

(9) $$\ln L = \sum_i NN_i \cdot \ln[\Pr(NN_i)] + NY_i \cdot \ln[\Pr(NN_i)] + YN_i \cdot \ln[\Pr(YN_i)] + YY_i \cdot \ln[\Pr(YY_i)].$$

Data

A sample of 733 completed responses was used for this analysis.[7] The sample was collected in the Fall of 2000 and consists of California residents selected at random with the assistance of a private sampling firm. Implementation of the mail survey was conducted following the principles of the Total Design Method (Dillman). Early versions of the survey were shown to individuals with knowledge and expertise in the areas of resource valuation and water quality engineering and policy. Four focus groups were then held to gauge reaction to question wording and content. Finally, a pre-test survey was mailed to randomly selected households in the community of Woodland in the Sacramento Valley, and the results and comments received were evaluated and incorporated into a final version. The overall response rate was 60.4% of deliverables, and 52% of the total initial mailing.

Table 1. Monthly Surcharge Threshold Levels

Variable	Symbol	Mean	Standard Deviation	Minimum	Maximum
Initial bid	B_I	\$41.19	\$12.41	\$20	\$60
High bid	B_H	\$73.08	\$13.44	\$50	\$90
Low bid	B_L	\$16.31	\$9.18	\$5	\$30

Open-ended responses from the focus groups and pre-test were used to determine the surcharge amounts used for the valuation questions in the final version. The monthly surcharge amounts offered in the final survey ranged from \$5 to \$90 corresponding to annual household surcharges of \$60 to \$1080. Table 1 provides a descriptive summary of the surcharges offered in the survey. As the table shows, the means of the lower, initial, and higher surcharges are \$16, \$41, and \$73, respectively.

Population characteristics of interest are presented in Table 2. The household income (INC) and years of educational attainment (EDUC) statistics were constructed from U.S. Bureau of Census data from the 1990 Census.[8] Household income was converted to current dollars using the Consumer Price Index. As mentioned before, the environmental concern population statistic was drawn from a 1997 California Field Poll

(Field Poll, 1997). The sample used in the statewide Field Poll survey was representative of California's population, having demographics that closely resembled the population's characteristics. The table shows that the mean California adult resident has an annual household income of $47,700, has completed high school, and is "somewhat" to "extremely" concerned about environmental issues. On the other hand, the median California adult resident lives in a household with an annual income of $42,100, has completed a year of school beyond high school, and is "extremely" concerned about environmental issues.

Table 2. Descriptive Statistics for the Sample

Variable	Household Income (INC)	Educational Attainment (EDUC)	Environmental Concern (CONC)
Sample Characteristics			
Units	$ per household	Years completed	Index
Sample Mean	$70,600	15.17	7.00
Sample Median	$62,500	15	7
Std Dev	$51.92	3.16	1.11
Min	$2,500	2	2
Max	$200,000	22	8
Sample Size	733	733	733
California Population Characteristics[a]			
Units	$ per household	Years completed	Index
Population Mean	$47,700	12.89	6.86
Population Median	$42,100	13	8

[a]Calculated from U.S. Bureau of Census data for 1990 Census, 1997 Census data, and 1997 California Field Poll results

Table 2 also contains some descriptive statistics of the data used in the analysis. The mean household income (INC) of the sample is about $24,000 more than the state average of $47,700 and about $29,000 more than the median state level of $42,100. The mean (median) number of years of education (EDUC) completed by the sample was 15.17 years (15 years), compared to a statewide mean (median) of 12.89 years (13 years). The environmental concern variable (CONC) is an index, representing the sum of responses to both attitudinal questions, with responses being assigned a value from a low of 1 to a high of 4. Thus, higher values for CONC represent more concern for pollution and the environment. The sample mean (median) value is 7.00 (7), so individuals in the sample averaged between a 3 and 4 on each question, suggesting a strong concern for the environment. This can be compared to the mean (median) of 6.86 (8) from the 1997 Field Poll (Field Institute, 1997), which was the most recent Field Poll that included the two environmental concern questions.

In addition, about eighty percent of the sample indicated they participate in recreation activities at or near water bodies in the state (rivers, lakes, estuaries, wetlands, and coastal waters).

Results

As noted, one of the purposes of this paper is to develop population WTP estimates based on the representativeness of the sample not only with respect to standard demographic variables, but also with respect to environmental attitude. Since higher numbers for CONC correspond to greater concern for environmental issues, we would expect WTP to increase with an increase in CONC. In addition to this attitudinal variable, INC and EDUC were included as covariates of WTP. Economic theory suggests that WTP for a "good" should increase with income, *ceteris paribus*, and income should therefore have a positive sign. In general, the marginal effect of education on WTP is likely to be dependent on the nature of the good; though for the case of improvements in water quality, it is likely that persons with higher educational attainment may have a better understanding of the benefits of water quality improvements. If this were the case, we would expect education to have a positive coefficient.

To evaluate the role of environmental attitude, two models were estimated, one with CONC as a covariate (Model A) and one without (Model B). The models were estimated by maximum likelihood techniques applied to equation (9), using GAUSS version 3.2.25, and the results compared.

Coefficient estimates for both models are reported in Table 3. In general, all estimated coefficients were statistically significant (based on asymptotic t-statistics) and positive, with the exception of the constants, which were negative. Additionally, the variance scale factors are estimated precisely, having asymptotic t-values greater than 6. For both models, the income parameters are statistically significant and of the expected signs (positive). For every additional $10,000 individuals earns per year, the models predict individuals would be willing to pay an additional amount in the range from $2.08 to $2.22 per month, or $24.96 to $26.64 per year. The models also predict education has a statistically significant positive marginal effect on the valuation of improving water quality, increasing the WTP by a little over $2 for each additional year of education completed.

Table 3. WTP Function Parameter Estimates for Models With and Without CONC

		MODEL A		MODEL B	
Parameter	***Variable***	***Estimate***	***Asymptotic Student's t***	***Estimate***	***Asymptotic Student's t***
$\beta_{Constant}$	Constant	-106.8897	-4.0057	-37.2600	-2.6201
β_{INC}	INC	2.218×10^{-4}	4.5265	2.081×10^{-4}	4.4657
β_{EDUC}	EDUC	2.0756	2.9626	2.1665	3.1376
β_{CONC}	CONC	9.8004	4.3246		
γ	Gamma	0.5806	2.5398	0.5220	2.4302
σ	Standard error	41.6218	6.6293	41.0644	6.8020
	Log-L	-784.67		-801.89	
	Log-L (β=0)	-1016.15		-1016.15	
	Pseudo-R^2	0.228		0.211	
	N	733		733	

The presence of anchoring is evident in both models, since γ is asymptotically statistically significant and different from zero in both sets of results. In Model A, the presence of the initial surcharge amount accounts for approximately 37 percent ($\gamma/(\gamma+1)$) of the observed WTP, while for Model B anchoring accounts for about 34 percent of observed WTP.

Greater concern for the environment, as indicated by responses to the Field Poll attitudinal questions, has a strong and statistically significant marginal effect in Model A, which included CONC as a covariate. The model predicts WTP increases by almost $10 with a unit increase in the concern index, which is equivalent, for instance, to going from "somewhat concerned" to "extremely concerned" in one of the two environmental concern questions.

The pseudo-R^2 reported in Table 3 is an informal goodness-of-fit measure analogous to the R^2 measure used in conventional regression models (Greene, 1993; Ben-Akiva and Lerman, 1985). It is defined as

$$\text{Pseudo-R}^2 = 1 - \frac{\ln L}{\ln L_0},$$

where $\ln L$ is the maximized log-likelihood function value for the full model and $\ln L_0$ is the maximized log-likelihood function value for the model with all coefficients set to zero. As Table 3 shows, the pseudo-R^2 for Model A is 0.228, and for Model B it is 0.211. Although commonly interpreted in a manner similar to R^2, Greene (1993) points out that models with a low pseudo-R^2 do not always indicate poor predictive power.

To evaluate whether CONC should be included as a covariate in the WTP function used to adjust WTP, a likelihood ratio test was performed. Using the log-likelihoods in Table 3, the χ^2 test statistic for the hypothesis that $\beta_{CONC} = 0$ is 34.44, which exceeds the $\chi^2_{.05,\ 1\ df}$ critical value of 6.63. The null hypothesis is rejected, and Model A is taken as the preferred model.

Because WTP is non-negative, a conditional censored WTP estimate is calculated.[9] Expected WTP is

$$E(WTP) = \Pr\left[\varepsilon_i \geq \frac{-x_i \cdot \beta}{\sigma}\right] \cdot E[WTP \mid WTP \geq 0] + \Pr\left[\varepsilon_i < \frac{-x_i \cdot \beta}{\sigma}\right] \cdot 0.$$
$$= \left(1 - \Phi\left(\frac{-x_i \cdot \beta}{\sigma}\right)\right) \cdot \left[x_i \cdot \beta + E\left[\varepsilon_i \mid \varepsilon_i \geq \frac{-x_i \cdot \beta}{\sigma}\right]\right]$$

since ε_i follows the standard normal distribution. Upon simplifying, mean WTP can be expressed as

$$E(WTP) = \left\{1 - \Phi\left(\frac{-x_i \cdot \beta}{\sigma}\right)\right\} \cdot \left[x_i \cdot \beta + \sigma \cdot \frac{\phi\left(\frac{-x_i \cdot \beta}{\sigma}\right)}{1 - \Phi\left(\frac{-x_i \cdot \beta}{\sigma}\right)}\right]. \tag{10}$$

Using (10) and the results from Table 3, Model A yields mean sample WTP of $23.01 per month, with a standard error of $4.74, calculated using the Krinsky-Robb simulation procedure. Mean WTP was also calculated for Model B for comparison, and the mean WTP was $23.06 per month, with a standard error of $4.40. Annually, this amounts to the average household paying $276.12 and $276.72 under models A and B, respectively. The sample mean WTP estimates from the two models are almost identical and not statistically different. Thus, while CONC itself is highly significant, its inclusion does not have a significant effect on WTP estimates.

Aggregate Benefits from Water Quality Improvement in California

To generate an estimate of WTP for the California population, the sample estimates must be adjusted to reflect characteristics of the general population. Model A is used for this purpose, as it was judged superior to model B on statistical grounds.

The WTP for the mean Californian is calculated by evaluating the estimated WTP function at the California mean levels of income, education, and environmental concern. While extrapolations based on the mean are usually preferred, we also calculate aggregate WTP by adjusting to the medians of each of the covariates.

The aggregate WTP estimates are displayed in Table 4. The monthly WTP based on means of the covariates is $15.46, while the monthly WTP evaluated at the medians household is $20.81, which is not statistically different (according to an asymptotic t-test). The higher California median values are attributable to the strong influence of the environmental concern variable.

Table 4. Adjusted Monthly WTP Evaluated at Mean and Median California Household Characteristics (Krinsky-Robb Standard Errors in Parentheses)

Model	Evaluated at Mean Characteristics of CA Households	WTP Evaluated at Median Characteristics of CA Households
WTP Estimate	$15.46 ($3.41)	$20.81 ($3.87)

According to the most recent Census data, there are about 10.4 million households in California. One way to calculate aggregate WTP for the level of water quality improvement being considered here is to assume all individuals in the state are identical to either the mean or median California household (e.g., Rubin, Helfand, and Loomis, 1987). Under this assumption, the aggregate WTP estimates are $1.9 billion and $2.6 billion per year, respectively (Table 5).

Table 5. Aggregate Annual California WTP Assuming Identical Households (in $millions)

Model	Mean CA Household Characteristics	Median CA Household Characteristics
Aggregate WTP Estimate	$1,930	$2,597

Discussion and Further Work

As public and private entities spend increasing amounts of money on pollution control and abatement activities, a natural question to ask is "What is the level of public benefits derived from these expenditures?" The primary purpose of this study was to attempt a first answer to that question. Using data from a contingent valuation survey, willingness to pay for restoring California water bodies to a water quality level where no impairments to beneficial uses remain was estimated as a function of demographic variables and environmental concern. The model results were corrected for anchoring effects caused by the presence of two valuation questions in the referendum with follow-up question format. Using population values of the willingness to pay covariates, the aggregate annual benefit of improving water quality to levels consistent with state and federal water laws and policy was estimated. The results suggest aggregate benefits of approximately $2 billion.

The use of attitudinal questions taken from readily-available surveys of the population of interest, such as National Gallup Polls and General Social Surveys, offers a potentially promising way for CV researchers to extrapolate sample results to the population. In this paper, we explored the use of an environmental concern variable constructed from responses to questions in the survey that are identical to questions found in a California general population poll. The environmental concern variable had a significant impact on the adjusted WTP since its coefficient was both strongly statistically significant and large enough in magnitude to have a strong marginal effect on WTP. This points to the need for further research on using attitudinal variables from general population surveys, such as the environmental concern variable used here, in contingent valuation studies that use willingness to pay functions to adjust sample results to the population.

References

Alberini, Anna, Barbara Kanninen, and Richard T. Carson. Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data. *Land Economics*, 73: 309-24, 1997.

Ben-Akiva, Moshe and Steven R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press, 1985.

Bockstael, Nancy E., Kenneth E. McConnell, and Ivar E. Strand. "Measuring the Benefits of Improvements in Water Quality: The Chesapeake Bay." *Marine Resource Economics*, 6: 1-18, 1989.

Cameron, Trudy 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, 1988.

Cameron, Trudy A. and John Quiggin. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-Up' Questionnaire." *Journal of Environmental Economics and Management,* 27: 218-234, 1994.

Carson, Richard T. and Robert C. Mitchell. "The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water." *Water Resources Research*, 29(7): 2445-2454, 1993.

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

The Field Institute, Field (California) Polls. 97-04. San Francisco, CA: The Field Institute. Berkeley, CA: University of California Data Archive, 1997.

Gramlich, Frederick W. "The Demand for Clean Water: The Case of the Charles River." *National Tax Journal*, 30(2): 183-194, 1977.

Greene, William H. *Econometric Analysis*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1993.

Hanemann, W. Michael, John Loomis, and Barbara Kanninen. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *American Journal of Agricultural Economics,* 73(4): 1255-1263, 1991.

Harrison, Glenn W. and James C. Lesley. "Must Contingent Valuation Surveys Cost So Much?" *Journal of Environmental Economics and Management*, 31: 79-95, 1996.

Herriges, Joseph A. and Jason F. Shogren. "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning." *Journal of Environmental Economics and Management*, 30: 112-131, 1996.

Krinsky, I., and A. L. Robb. On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics*, 68: 715-719, 1986.

Larson, Douglas M., Daniel K. Lew, and Yuko Onozaka. "Clean Water in California: What Do You Think? Preliminary Results from the California Water Quality Improvement Survey." Department of Agricultural and Resource Economics, University of California, Davis, February 2001.
Loomis, John B. "Expanding Contingent Valuation Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions." *Land Economics*, 63(4): 396-402, 1987.
McLeod, Donald M. and Olvar Bergland. "Willingness-to-Pay Estimates Using the Double-Bounded Dichotomous-Choice Contingent Valuation Format: A Test for Validity and Precision in a Bayesian Framework." *Land Economics*, 75(1): 115-125, 1999.
Mitchell, Robert C. and Richard T. Carson. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington, D.C.: Resources for the Future, 1989.
Oster, Sharon. "Survey Results on the Benefits of Water Pollution Abatement in the Merrimack River Basin." *Water Resources Research*, 13(6): 882-884, 1977.
Rubin, Jonathan, Gloria Helfand, and John Loomis. "A Benefit-Cost Analysis of the Northern Spotted Owl." *Journal of Forestry*, 89(12): 25-30, 1991.
Stanley Hoffman & Associates. "Financial and Economic Impacts of Storm Water Treatment Statewide NPDES." Report to the California Department of Transportation Environmental Program, 1998.
State Water Resources Control Board. *1998 California Water Quality Assessment* Staff report, Division of Water Quality, SWRCB, August 1999.
U.S. Environmental Protection Agency. *The Quality of Our Nation's Waters: A Summary of the National Water Quality Inventory 1998 Report to Congress.* Office of Water, Washington D.C., Report EPA 841-S-00-001, 2000.

Endnotes

[1] "Impaired" waters do not support one or more of the beneficial uses designated for them by the State Water Resources Control Board. Pollutant levels in these waters exceed the water quality objectives set forth for protecting beneficial uses.

[2] Water bodies included were lakes, rivers, streams, estuaries, wetlands, and coastal waters. Groundwater was not included.

[3] The sampling frame of the California Field Poll is California households as was for our survey. The 1997 poll included responses by 1006 people, whose mean characteristics (age, income, education, and household size) are closely aligned to that of the state population.

[4] Another common distributional assumption is that ε_i follows a logistic distribution, leading to logit estimation.

[5] With mail surveys, there is an additional concern that referendum questions with follow-ups may introduce bias to responses since respondent can view all questions at once, similar to a payment card format. Thanks to John Loomis for pointing this out.

[6] NN is an indicator variable that takes a value of 1 for cases where the individual says "no" to both the initial and follow-up surcharge amounts. NY, YN, and YY are similarly defined.

[7] For a detailed description of the data, see Larson, Lew, and Onozaka (2001).

[8] 2000 Census data was not yet available at the time this analysis was conducted, but will be used in subsequent analyses as it becomes available.

[9] A similar approach was taken by McLeod and Bergland (1999) for calculating conditional WTP.

Measuring Individual-Specific Shadow Values of Time

Douglas M. Larson

And

Daniel K. Lew

Department of Agricultural and Resource Economics
University of California, Davis.

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Abstract: Empirical measurement of the shadow values of leisure time, or shadow wages, is important for work in several areas, including the valuation of natural amenities through associated recreation demand, evaluating transportation projects, valuing home production, and in litigation over loss of life or limb. The methods that that have come into use for this purpose typically do not differentiate between the types of non-work time when measuring shadow values. Yet it is recognized that people allocate their non-work time subject to a variety of constraints on "the timing of leisure time;" i.e., differential availability of the opportunity to participate in non-work activities.

This paper suggests a utility-theoretic, survey-based approach for measuring an individual's multiple shadow values for different non-work activities. The individual's relative preferences for activities, along with their time and money prices, "reveal" the shadow values directly. Hypothesis testing for differences among the shadow values for different activities allows the researcher to identify a minimal set of time constraints that reflect the individual's stated preferences and choices. An empirical application illustrates the ease of use of the approach.

Developing better empirical measures of the shadow value of time is important in several areas of applied economics, including the evaluation of transportation projects (de Donnea; Quarmby), adjusting the national accounts for the value of home production (Gronau; Hersch), assessing the value of natural resources that support outdoor recreational activity (Knetsch; Smith *et al.*; Johnson), and understanding labor market choices both in developed and developing economies (Rosenzweig; Strauss; van Soest).

A number of approaches have been developed to estimate shadow values in the literatures on labor supply (e.g., Heckman; Wales and Woodland; Zabel; Macurdy *et al.*), farm household consumption and production choices (e.g., Jacobi; Lopez), and recreation demand (Feather and Shaw). Typically, these models are motivated by a consumer or producer making choices subject to a single constraint on time, in addition to money budget or technology constraints, and therefore yield a single estimate of the shadow, or scarcity, value of time. Yet, as Smith *et al.* have pointed out, this is a fairly simple treatment of time as a constraint. Often, the "timing of the time" is also important, as some activities are conducted during periods when time constraints bind much more tightly than in other periods, as anyone who has rushed to finish writing an exam before rushing off to get on an airplane can attest.

This paper aims to do two things. First, it develops a theoretical framework for interpreting consumer choices made subject to multiple binding time constraints. One of the implications of such a framework is that different activities have different scarcity values of time, depending on when during the course of the day, week, or month they take place. The second, and more novel, contribution is to show how one can measure the multiple scarcity values of time that an individual may experience. Using surveys that ask people how they like different activities at the margin, and collect information on how they allocate their time among activities with different money prices of consumption, the resulting data enable the estimation of the scarcity values of time that must underlie the individual's "observed" uses of her time.

There are two important features of the approach. First is that one can use the data supplied by an individual to estimate that person's shadow values, avoiding the need for interpersonal comparisons that would require full cardinality of utility. Instead, by using intra-person information on relative strength of preference for activities, the shadow values that result require only ordinal representations of the underlying utility function, as they are invariant to monotonic increasing transformations of utility.

A second important feature of the approach is that it is not necessary to ask the individual how many different constraints they face, which would be a difficult question to answer realistically. Instead, one can use the fact that activities that are chosen from the same time constraint have the same shadow value of time to determine empirically how many different shadow values explain the reported choices. Beginning with a "naive" model that allows (nearly) each activity to have its own shadow value, by sequentially testing equality restrictions on the shadow values, one can arrive at a specification of the minimum number of unique (stastically different) shadow values that the data support. Even though the number of degrees of freedom are small, the precision of shadow value estimates is sufficiently high to enable rejection of the equality restrictions when the number of shadow values is relatively small.

The first section develops the model of consumer choice subject to multiple binding time constraints. Then the empirical estimation approach is illustrated for two individuals using data from a pre-test of the survey. The paper closes with some remarks about the approach and the results obtained, along with extensions that could be implemented.

Consumer Choice With Multiple Time Constraints

To develop the conceptual model, consider a consumer with utility function u(**x**), where **x** is an n-vector of activities or (used interchangeably) consumption goods. Activities have money prices and time requirements for consumption, so that the consumer is constrained by both time and money in making her choices. Because different activities might have different shadow values, the individual's overall time T is assumed to be representable by a series of mutually exclusive and exhaustive time constraints T^k, so that

$$T = \Sigma_k T^k, \tag{1}$$

where k=1,...,N indexes the different periods that T is comprised of. The consumer does not have the ability to adjust the individual time constraints, so the marginal utilities of the T^k's are, in general, not equal. This gives rise to different shadow values of activities, depending on which time budget they are taken out of. Also, because overall time is partitioned into the fixed periods $T^1,...,T^N$, the marginal utilities of some types of time, that are perhaps necessary but not enjoyable, can be negative. Examples might include doing housework, visiting the in-laws, or activities undertaken out of a sense of duty rather than for their own pleasure.

Each activity x_i^k (with $i=1,\ldots,n_k$) is assumed to have a money price $p_i^k \geq 0$, and a time "price" $t_i^k \geq 0$. The time price formulation recognizes that most types of consumption require time in addition to money in consumption. A common example is outdoor recreation, where some time must be spent in transit to the place of consumption, in addition to the time spent in the outdoor recreation activity itself. In fact, most activities (e.g., eating, shopping, reading, watching movies) require at least some time in consumption, though whether this is a major part of the cost to the consumer depends on the type of good, the technology of consumption, and the consumer herself. The time price, for an individual, is taken to be fixed and exogenous, like the money price.

As special cases, a good could have either a zero time price in consumption or a zero money price in consumption, but not both. This recognizes that consumption will always be costly, but accommodates a wide range of activities that have only one type of price as well as those that have both. For time spent earning money (i.e., in labor), the price is a net price representing the difference between outlay (e.g., for meals and snacks) and the wage earned per hour, which will typically be negative. For activities that do not generate income, the price will be non-negative (though typically positive).

An aside on the definition of time "prices" may be helpful. How they are defined is necessarily a consequence of how the activity being valued is defined. Using the outdoor recreation example, some travel from one's home to a distant recreation site is required to consume the activity. Suppose it takes an hour each way to travel to a state park, and the recreationist spends 4 hours at the park. By defining the activity being valued (i.e., the activity that generates utility or disutility) as consumption on-site, the time price is 6/4=1.5, reflecting the fact that 1.5 hours of total time are required to consume each hour of the activity.

Alternatively, one could define the activity being valued as time away from home. In this case, all time spent is part of consumption, as opposed to being simply part of the price paid to gain access to consumption. The time price under this definition is 6/6=1, as all six hours away from home contribute to the activity being consumed.

Other activities conducted away from home similarly have fixed and variable time requirements in consumption, and whether the time price is 1 or greater than 1 depends on whether both expenditures of time are considered part of the activity, or whether the access time is viewed as simply part of the cost. In many cases, it is reasonable to expect that time prices are 1 by virtue of the activities being valued, but the framework accommodates other cases where it is not appropriate to assume time prices are 1.

The formulation of the time constraint in (1) also assumes that each activity occurs only within a single time constraint.[1] This simplification does not restrict the generality of the overall approach, which could easily be reformulated to accommodate activities appearing in multiple constraints.[2]

The consumer's choice problem is then

$$\text{Max}_{\mathbf{x}}\ u(\mathbf{x}) + \lambda[M - \mathbf{px}] + \Sigma_k\, \mu_k[\, T^k - \mathbf{t}^k\mathbf{x}^k] \tag{2}$$

where $\mathbf{x}^k$, k=1,...,N, are the groups of activities appearing in time constraint T^k, and $\mathbf{x} = [\mathbf{x}^1,...,\mathbf{x}^k,...,\mathbf{x}^N] = [x_1^1, \ldots, x_{n_1}^1, \ldots, x_1^N \ldots, x_{n_N}^N]$ is the full consumption vector whose corresponding price vector is $\mathbf{p} = [\mathbf{p}^1,...,\mathbf{p}^k,...,\mathbf{p}^N] = [p_1^1, \ldots, p_{n_1}^1, \ldots, p_1^N \ldots, p_{n_N}^N]$. The money constraint is assumed to be binding, so that the conditions for optimal choice of the activities an individual participates in (i.e., for $x_i^k \geq 0$) are

$$u_i^k - \lambda\, p_i^k - \mu_k t_i^k = 0, \qquad \lambda > 0, \qquad \mu_k \gtreqless 0, \tag{3}$$

for $i = 1,...,n_i$, $k = 1,...,N$.

Dividing (3) by the marginal utility of money, λ, and rearranging terms, one obtains the familiar equality of marginal value and marginal cost of goods that are consumed,

$$u_i^k/\lambda = p_i^k + (\mu_k/\lambda)\cdot t_i^k, \tag{4}$$

or

$$MV_i^k = p_i^k + \nu_k \cdot t_i^k, \tag{5}$$

where $MV_i^k \equiv u_i^k/\lambda$ is the marginal value of good x_i^k and $\nu_k \equiv \mu_k/\lambda$ is the scarcity value of time in constraint k. In particular, ν_k is the shadow value of time for good x_i^k and all other goods in the k^{th} time constraint.

An important point to note is that the shadow values MV_i^k and ν_k in equation (5) are invariant to monotonic increasing transformations of the utility function. Thus they are based on ordinal, not cardinal, utility. This can be seen by replacing the utility function $u(\mathbf{x})$ with the increasing transformation $T[u(\mathbf{x})]$. The problem is now

$$\text{Max}_{\mathbf{x}}\ T[u(\mathbf{x})] + \lambda'[M - \mathbf{px}] + \Sigma_k\, \mu'_k[\, T^k - \mathbf{t}^k\mathbf{x}^k] \tag{2'}$$

where the shadow values λ' and μ'_k are in general different from the original λ and μ_k. The new first order conditions are

$$T' \cdot u_i^k - \lambda' p_i^k - \mu_k' \cdot t_i^k = 0. \tag{3'}$$

where $T' \equiv \partial T/\partial u$. The same optimal **x** solve (2) and (2′), since monotonic increasing transformations of the utility function don't affect observed consumption choices. This means u_i^k is the same in (3) and (3′), and dividing through (3′) by T′ and comparing with (3), it is apparent that the relationships between the shadow values λ', μ_k', λ, and μ_k are

$$\lambda = \lambda'/T' \quad \text{and} \qquad \mu_k = \mu_k'/T'.$$

Replacing λ' with $\lambda T'$ and μ_k' with $\mu_k T'$ in (3) and dividing by T′, it can be seen that (3′) is identical to (3), so the same shadow values MV_i^k and ν_k result under the transformed utility function.

Estimating Individual-Specific Scarcity Values of Time

The individual's relative preference for good x_i^k over good x_j^h, S_{ikjh}, can be expressed as the ratio of the marginal utilities of the two goods, $S_{ikjh} = u_i^k/u_j^h$, or as the ratio of their marginal values,

$$S_{ikjh} = (u_i^k/\lambda)/(u_j^h/\lambda).$$

In light of (4) and (5), this can also be written as

$$S_{ikjh} = (p_i^k + \nu_k \cdot t_i^k)/(p_j^h + \nu_j \cdot t_j^h). \tag{6}$$

Equation (6) is the key equation used in the empirical application to assess the opportunity costs of time associated with different activities. It is a direct extension to the multiple-time constraint setting of the usual equality of marginal rates of substitution between two goods to their relative price ratios, where the relevant prices here are full prices that are dependent on endogenous shadow values of time.

The left hand side of (6), S_{ikjh}, is the relative strength of preference for the activities x_i^k and x_j^h, i.e., the ratio of their marginal utilities. Whereas in the single- (money) constraint case the price ratio reveals S_{ikjh} directly, in (6) there are two unknowns on the right side, namely the scarcity values of time for the two activities.[1] Thus, in the multiple- time constraint case, (6) does not reveal S_{ikjh} unless ν_k and ν_j are known. Alternatively, though, if the marginal rate of substitution between activity i and j is known, equation (6) can be used to estimate ν_k and ν_j.

The relative marginal utilities of activities can, in principle, be obtained through ratings surveys that are an increasingly-familiar cognitive exercise for today's consumers. Ratings are a common and widely understood mechanism for conveying relative marginal utility (and sometimes marginal value), as evidenced by the popularity of consumer guide books such as *Consumer Reports*. By asking a consumer to rate different activities according to the marginal utility they provide, the ratios of those ratings convey the relative preferences required to construct S_{ikjh}. Provided that the ratings scale is sufficiently flexible to accommodate a wide range of relative ratings, the induced variable S_{ikjh} can be treated as continuous for estimation purposes.

By appending an additive error to (6), the relative preferences can be written as

$$S_{ikjh} = (p_i^k + \nu_k \cdot t_i^k)/(p_j^h + \nu_j \cdot t_j^h) + u_{ij}, \qquad i=1,\ldots,n_k, j=1,\ldots,n_h\,,\ i \neq j \qquad (7)$$

If data are also collected on the money and time prices the individual faces for each activity, equations (7) represent a system of equations for the individual from which that person's shadow values ν_i can be estimated, via nonlinear least squares or maximum likelihood. An individual's rankings of n activities provides n(n-1) observations on relative ratings S of pairs of activities, though only n-1 are unique. Thus up to n_m-1 individual shadow values can be identified from information on n_m activities provided by a given individual m.

A key question is how many shadow values should be estimated. This question arises immediately upon noticing that at most, only n_m-1 shadow values can be estimated from information on n_m activities. This implies that at least one equality restriction is necessary to obtain parameter estimates.

More fundamentally, though, the issue turns on how many constraints the individual faces in the activities of interest to the researcher, and which constraints pertain to which activities. Both of these are unknown. One could attempt to ask the individual directly for this information, but it is unlikely that it could be obtained reliably or accurately.

This paper uses a sequential hypothesis testing procedure to determine the number of unique scarcity values applicable to the activities of interest in the analysis. The estimation strategy relies on the fact that activities that are chosen from the same time constraint have the same scarcity value.

Initially, each activity but one is allowed to have a unique scarcity value of time. Hypothesis tests of equality of estimated scarcity values are performed to determine whether the set of shadow values can be reduced. The scarcity values chosen for hypothesis testing in each step are those which have the smallest likelihood ratio or pairwise Student's-t statistic on the null hypothesis of coefficient equality. The testing sequence stops when hypothesis tests on equality of all remaining shadow values reject the hypothesis of equality. At this point, the remaining scarcity values, applicable for groups of activities, are significantly different from zero, and from each other.

Data

The data required for the estimation of time shadow values are ratings of satisfaction or marginal utility of different activities that are consumed, and their time and money prices. Results based on responses by two individuals to a survey on how they spent their time, and how they enjoyed the activities they spent time on, are developed and presented. This serves to illustrate both the hypothesis testing procedure, and that the estimation strategy produces individual-specific shadow values of time.

Both respondents were asked about a set of 18 different activities in which they may have participated during the previous week, under the broad categories of household work, school work, employment, and leisure time activities.

Under household work, the activities mentioned were washing the dishes, washing clothes, cleaning the house, and cleaning the yard. In the school work category, the activities asked about were attending lectures, attending discussion section or lab, studying, and travelling to and from school. Employment-related activities were time spent at the workplace and spent traveling to and from work. The leisure-time activities included reading for pleasure, watching

TV, playing on the computer, eating meals at home, eating meals out, going to a movie, playing sports, and exercising or working out.

Each respondent was asked to indicate how many hours per week s/he had spent in each activity, and what the total money cost for the week was. They were instructed to include the variable costs that they incur less frequently than on a weekly basis, such as monthly health club fees or purchase of laundry soap, prorated to a weekly cost. They were asked not to include durables such as the cost of washing machines to do the laundry. Not all activities have money prices (e.g., some forms of working out or cleaning the yard), so zero can be an appropriate money price, depending on the activity. The price applicable to work time is typically negative because the wage enters with a negative sign, as noted earlier. All of the time spent in each activity was assumed to be utility-generating, so that the time prices are 1, following the earlier discussion of defining time prices.

The respondents were also asked to rate, on a Likert scale from 1 to 10, how they liked each activity that they participated in. The instructions asked that they consider how much they liked the marginal unit of consumption, by asking them to consider how much they'd like a little more or less time spent in each activity per week, to help obtain marginal utility ratings. The Likert ratings can be viewed as scaled marginal utilities, and provided the rating-marginal utility correspondence is affine, the ratios of Likert ratings are marginal rates of substitution or ratios of marginal utilities.

The number of activities that a person had actually participated in determines n_m , the number of possible shadow values. Respondent 1 was a graduate student, while respondent 2 was a non-student, so respondent 1 participated in the 4 schoolwork activities while 2 did not. Overall, respondent 1 participated in 14 of the 18 possible activities while 2 participated in 12 (Table 1).

Results

For each individual, ratios of the Likert ratings S for each activity were formed, and these along with the money prices of each activity were used in estimating the system of equations

$$S_{ikjh} = (p_i^k + \nu_k)/(p_j^h + \nu_j \cdot t_j^h) + u_{ij}, \qquad i=1,\ldots,n, j=1,\ldots,n, i \neq j \qquad (8)$$

Equations (8) were estimated by nonlinear least squares, using Gauss version 3.2.25.

Table 2 shows the sequence of model estimation and hypothesis testing for person 1. Initially, each activity, 1-14, has its own scarcity value, ν_1-ν_{14}, with the activity that corresponds to work time (activity 8 in the case of person 1, activity 4 for person 2) labeled as ν_w. In each model, beginning with the first, successive equality restrictions between parameters are imposed, based on which pair had the lowest χ^2 statistic for the equality restriction. The restriction imposed can be identified by the scarcity value in **bold** type, which indicates what the activity scarcity value was set to. The χ^2 statistic for the test is reported at the bottom of each column. Coefficient estimates and χ^2 statistics in **bold** type are statistically different from zero at the 5% significance level.

Not surprisingly, equality restrictions among estimated scarcity value parameters are not rejected for the first several hypothesis tests, because degrees of freedom are low and standard errors of the parameter estimates are high. For person 1, it is not until Model 6, when eight scarcity values are estimated and 6 equality restrictions are imposed, that the χ^2 statistic rejects an equality restriction (at the 5% level, though not at the 1% level). At this juncture, the model has more than the optimal set of scarcity values to adequately describe person 1's relative preferences, money prices, and activity levels: only two of the six scarcity values are statistically different from zero, the coefficient on work time, with a scarcity value of \$21.40/hour, and the scarcity value on washing clothes, with a scarcity value of -\$4.92/hr.

A smaller set of scarcity values that more completely describe the activity set may be preferable, so further equality restrictions are tested, resulting in Model 9. In Model 9, five scarcity values are estimated, and all are significantly different from zero and from each other, at the 5% level. Model 9 is the preferred model, as no equality restrictions were rejected in moving from Model 5 to Model 9, and when one imposes additional restrictions beyond Model 9, they are strongly rejected by the data (e.g., a χ^2 statistic in imposing an additional restriction on Model 9 of 26.3, significant at the 0.5% level). It is also a more parsimonious model in that fewer parameters (five) are estimated, with greater precision.

Table 3 presents more detail on both models, providing the estimated scarcity values and their Student's-t statistics for a test of difference from zero. In the preferred Model 9 for Person 1, five activities had negative scarcity values, including 4 (eating meals at home, eating out, washing the dishes, and cleaning house) with a scarcity value of -\$1.04/hour, and 1 (washing clothes) with a scarcity value of -\$5.04/hr. The remaining 9 had positive scarcity values, including 6 (all four school-related activities, reading for pleasure, and playing on the computer) with a scarcity value of \$8.31/hr., work time with a scarcity value of \$20.78/hr., and the leisure activities of working out and going to a movie, with scarcity values of \$5.43/hr.

Because time constraints are generally strictly binding—time must be "spent" in one way or another—there is nothing intrinsically surprising about negative scarcity values of time. They indicate blocks of time that the individual would prefer were shorter in duration, but cannot be made so because of the inability to fully rearrange time between uses. This may be because the activities chosen within the block of time themselves are not enjoyable, but are necessary for the longer-term well-being. One might imagine going to the dentist as such an activity, whose marginal value is negative. Because, according to equation (5), the (negative) marginal value equals the (positive) money price plus the scarcity value of time for the activity, this would assure that the scarcity value associated with going to the dentist was negative.

Alternatively, negative scarcity values may arise when an activity is enjoyable, but simply takes too long. In such cases, the marginal value of an activity within the constraint is positive, but because the money price is greater, the scarcity value must be negative when the activity is observed to be undertaken. Examples no doubt would vary greatly based on individual preference, but examples might include going clothes shopping, or travel to distant locations to visit relatives or enjoy recreational activity.

Table 4 presents person 1's marginal values for each activity, along with their scarcity values, for Models 5 and 9. (The differences between the two are the monetary prices per hour.) All marginal values are positive, ranging from just under \$1/hr to \$9.30/hr, indicating that Person 1 enjoys all activities. However, for the cleaning (1-3) and eating (11-12) activities, the marginal value per hour is not as high as the monetary cost per hour, implying a negative scarcity value of time for those activities.

Person 1, the student, has a wage of \$16/hour, which shows up as a price of -\$16/hr for work time (activity 8). The scarcity value of work time is \$20.78, so that the s/he has a marginal value of work time of \$4.78. Because this person enjoys work time, the wage does not fully cover the opportunity cost of work time.

Tables 5-7 go through a similar analysis of scarcity values and marginal values for Person 2, the non-student. In the case of person 2, the magnitudes of the scarcity values were quite similar for washing dishes and clothes (activities 1-2) and all leisure time activities except eating meals out (activity 9), throughout the analysis (Table 5). Starting from Model 3, where the degrees of freedom in estimation reduced the critical values of the Student's-t test to 4 and below, all but one of the scarcity values were significantly different from zero, though not from each other. The similarity of the scarcity values suggests that these activities were chosen from within a common time constraint, and hypothesis testing essentially confirmed this. Equality restrictions placed on estimated scarcity values were not rejected until Model 10, which had only two scarcity values explaining all the data and performed much worse than Model 9. Model 9 is the model which best explains the data, in terms of a scarcity value for work time (-$2.68/hr), a scarcity value for all leisure and housework except cleaning house ($14.94/hr), and a scarcity value for cleaning house ($6.45/hr). All scarcity values are significantly different from each other and from zero (Table 6). The negative scarcity value of work time suggests that if this person worked less s/he would be better off (that is, if the time constraint within which work time is chosen had less hours).

Table 7 presents the estimated scarcity values and the implied marginal values of activities for person 2. All marginal values are positive except for work time, and most are similar in magnitude, ranging from $15-$18 per hour, except for cleaning house and travel time to work (about $7/hour) and work time (-$20/hour). The negative marginal value of work is an example of how disliking an activity can give rise to a negative scarcity value. In this case, person 2's wage is $17.75/hr, but this does not fully compensate for the marginal disutility of work of -$20.43/hr.

Conclusions, Limitations, and Extensions

This paper has developed a simple empirical model to estimate the scarcity values of different activities a person participates in, based on the theory of consumer choice subject to multiple binding time constraints. The approach uses information on a person's relative preferences for different activities, along with the money and time prices and consumption levels of each, to estimate individual-specific shadow values of the activities. These shadow values appear as part of the full prices in the expressions of equality of marginal rates of substitution between activities to their full price ratios.

The modeling approach recognizes that "the timing of time" is important; that is, that multiple time constraints may bind in different ways at different times and for different activities for an individual, thereby generating different shadow values representing the opportunity costs of those activities. These opportunity or scarcity costs are important in a variety of applications, both in research and in private enterprise. Being able to estimate them them in a rigorous hypothesis testing framework, using readily obtainable data and easily applied methods, is important.

Some of the advantages of the modeling approach include the fact that it avoids interpersonal comparisons of utility, instead relying on information provided by an individual to estimate that person's shadow values of different activities. The shadow value estimates are invariant to the form of the underlying utility function, so they are based on ordinal utility functions. The ratings of activities that are used in estimation are a familiar exercise cognitively and can be obtained via simple survey research techniques. Perhaps most importantly, it is not necessary to specify in advance how many time constraints the consumer faces, nor which activities are chosen within which constraints. The process of testing for significant differences in shadow values determines how many shadow values are needed to adequately represent the data.

This empirical model was applied to data collected from two different individuals, one of whom provided information on 14 activities that s/he had participated in during the previous week; the other had participated in 12 activities. For individual 1, the preferred model that emerged from sequential hypothesis testing had 5 unique shadow values, all of which were significantly different from each other and from zero. For individual 2, the preferred model had 3 unique shadow values, also significantly different from each other and from zero. The fact that the preferred models for each individual had multiple statistically different scarcity values strongly suggests that models which estimate only a single scarcity value of time do not adequately reflect the differences in the costliness of time, depending on which activity is considered.

Negative scarcity values can occur in this empirical model, and are a reflection of the fact that time constraints are strictly binding, and often there is limited ability to reschedule activities between time constraints. They may also reflect the fact that some activities that are not enjoyed must nevertheless be undertaken. They indicate that the individual could benefit were it possible to make some activities and uses of time shorter.

For individual 1, two of the five estimated scarcity values, those pertaining to cleaning and eating, were negative, though all estimated marginal values were positive. This implies that individual 1 enjoyed all activities, but some were not enjoyed as much as their costs of consumption. For individual 2, only work time had a negative scarcity value, and it also had a negative marginal value.

It is important also to note some potential limitations of the results presented here. The marginal and scarcity values may be somewhat sensitive to the fineness of gradation of the ratings scale used to collect relative preference data. While a 10-point Likert scale was used in this application, it may be that finer resolutions of the preference scale will produce more precise shadow value estimates. While ratings are a cognitively familiar exercise, it is important that ratings obtained be for marginal utilities rather than total or average utilities. Collecting money price information about different activities can be complicated by the fact that people do not always think of what they are spending in activities for which purchases are infrequent or irregular. Focusing on variable costs of consumption is appropriate for generating short-term scarcity value estimates, but for life cycle applications it may be necessary to consider the role of durables purchases for at least some activities.

Footnotes

1. For simplicity, consumption is also assumed to be non-joint; that is, within each time constraint, activities are mutually exclusive and exhaustive.
2. If consumption of an activity were positive in two different time constraints, either the shadow values of the constraints must be equal (in which case they could be combined into a single constraint) or the activity must have different marginal utility or money price in each constraint. In the latter case, the activity could be considered as two separate activities whose (presumably different) scarcity values could be estimated separately.
3. While these scarcity values could be equal, in general this needn't be the case.

References

de Donnea, F. X. "Consumer Behavior, Transport Mode Choice and Value of Time: Some Micro-Economic Models." *Regional and Urban Economics* 1 (1972): 355-382.

Gronau, R. "Leisure, Home Production, and Work--the Theory of the Allocation of Time Revisited." *Journal of Political Economy* 85 (1977): 1099-1123.

Grossbard-Schechtman, S., and S. Neuman. "Labor Supply and Marital Choice." *Journal of Political Economy* 96 (1988): 1294-1302.

Hersch, J. "The Economics of Home Production." *Southern California Review of Law and Women's Studies* 6 (1997): 421-440.

Heckman, J. J. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42 (1974): 679-694.

Jacobi, H. "Shadow Wages and Peasant Family Labor Supply: An Econometric Application to the Peruvian Sierra." *Review of Economic Studies* 60 (October 1993): 903-921.

Johnson, M. B. "Travel Time and the Price of Leisure." *Western Economic Journal* 4 (Spring 1966): 135-145.

Knetsch, J. L. "Outdoor Recreation Demands and Benefits." *Land Economics* 39 (1963): 387-96.

Macurdy, T., D. Green, and H. Paarsch. "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply." *Journal of Human Resources* 25 (1990): 415-490.

Lopez, R. "Estimating Labor Supply and Production Decisions of Self-Employed Farm Producers." *European Economic Review* 24 (1984): 61-82.

Moffitt, R. "The Estimation of a Joint Wage Hours Labor Supply Model." *Journal of Labor Economics* 8 (1990): 550-566.

Quarmby, D. A. "Choice of Travel Mode for the Journey to Work." *Journal of Transport Economics and Policy* 1 (1967): 273-314.

Rosenzweig, M. R. "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing Country." *Quarterly Journal of Economics* 95 (1980): 31-55.

Smith, V. K., W. H. Desvousges, and M. P. McGivney. "The Opportunity Cost of Travel Time in Recreation Demand Models." *Land Economics* 59 (1983): 259-277.

Strauss, J. "The Theory and Comparative Statics of Agricultural Household Models: A General Approach." Appendix in I. J. Singh, L. Squire, and J. Strauss, eds., *Agricultural Household Models: Extensions, Applications, and Policy.* Baltimore: Johns Hopkins University Press, 1986.

van Soest, A. "Structural Models of Family Labor Supply." *Journal of Human Resources* 30 (1995): 63-88.

Wales, T. and A. Woodland. "Estimation of Household Utility Functions and Labour Supply Response." *International Economic Review* 17 (1976): 397-410.

Zabel, J. "The Relationship Between Hours of Work and Labor Force Participation in Four Models of Labor Supply Behavior." *Journal of Labor Economics* 11 (1993): 387-416.

Table 1. Activity Lists

Activity	Person 1's Activity Number	Person 2's Activity Number
Household Work		
Washing the dishes	1	1
Washing clothes	2	2
Cleaning the house	3	3
Cleaning the yard		
School Work		
Attending lectures	4	
Attending discussion section/lab	5	
Studying	6	
Traveling to and from school	7	
Employment		
Time spent at the workplace	8	4
Traveling to and from Work		5
Leisure Time Activities		
Reading for pleasure	9	6
Watching TV		7
Playing on computer	10	
Eating meals at home	11	8
Eating meals out	12	9
Going to a movie	13	10
Playing sports		11
Working out/exercising	14	12

Table 2. Sequential Testing for the Shadow Values of Person 1

Model	1		2		3		4		5	
Activity	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate
1	ν_1	-1.0122	ν_1	-1.0517	ν_1	-0.99	ν_1	-1.1524	ν_1	**-1.3175**
2	ν_2	-5.0122	ν_2	-5.0517	ν_2	**-4.99**	ν_2	**-5.1524**	ν_2	**-5.3175**
3	ν_1	.	ν_1	.	ν_1	.	ν_1	.	ν_1	.
4	ν_4	8.8901	ν_4	8.8778	ν_4	9.095	ν_4	7.5524	ν_6	.
5	ν_5	9.8779	ν_5	10	ν_5	10.1	ν_5	8.4762	ν_5	6.8248
6	ν_6	7.8023	ν_6	7.9	ν_6	8.03	ν_6	6.7921	ν_6	5.643
7	ν_7	7.9023	ν_7	8.	ν_6	.	ν_6	.	ν_6	.
8	ν_W	20.9389	ν_W	21.	ν_W	**21.05**	ν_W	**20.2381**	ν_W	**19.4124**
9	ν_9	8.8779	ν_4	.	ν_4	.	ν_4	.	ν_6	.
10	ν_{10}	8.1758	ν_{10}	8.2857	ν_{10}	8.3757	ν_6	.	ν_6	.
11	ν_{11}	0.2124	ν_{11}	0.2857	ν_{11}	0.3457	ν_{11}	-0.6286	ν_{11}	-1.6194
12	ν_{12}	-1.1099	ν_{12}	-1.	ν_{12}	-0.91	ν_{12}	-2.3714	ν_{12}	-3.8576
13	ν_{13}	5.3901	ν_{13}	5.5	ν_{13}	5.59	ν_{13}	4.1286	ν_{13}	2.6424
14	ν_{14}	5.9267	ν_{14}	6.	ν_{14}	6.06	ν_{14}	5.0857	ν_{14}	4.0949
χ^2		-1.1E-12		5.62E-13		.0099		.0987		1.79

Model	6		7		8		9		10	
Activity	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate
1	ν_1	-0.9204	ν_1	**-0.8461**	ν_1	**-0.7057**	ν_{11}	.	ν_{11}	.
2	ν_2	**-4.9204**	ν_2	**-4.8461**	ν_2	**-4.7057**	ν_2	**-5.0432**	ν_2	**-4.9699**
3	ν_1	.	ν_1	.	ν_1	.	ν_{11}	.	ν_{11}	.
4	ν_6	.	ν_6	.	ν_6	.	ν_6	.	ν_6	.
5	ν_6	.	ν_6	.	ν_6	.	ν_6	.	ν_6	.
6	ν_6	9.4136	ν_6	**10.0829**	ν_6	**11.3462**	ν_6	**8.3092**	ν_6	**8.2202**
7	ν_6	.	ν_6	.	ν_6	.	ν_6	.	ν_6	.
8	ν_W	**21.3978**	ν_W	**21.7696**	ν_W	**22.4714**	ν_W	**20.7842**	ν_W	**21.1505**
9	ν_6	.	ν_6	.	ν_6	.	ν_6	.	ν_6	.
10	ν_6	.	ν_6	.	ν_6	.	ν_6	.	ν_6	.
11	ν_{11}	0.763	ν_{11}	1.2092	ν_{11}	1.85	ν_{11}	**-1.0432**	ν_{11}	**-0.9699**
12	ν_{12}	-0.284	ν_{12}	0.3853	ν_{11}	.	ν_{11}	.	ν_{11}	.
13	ν_{13}	6.216	ν_{13}	**6.9044**	ν_{13}	**7.9571**	ν_{13}	**5.4263**	ν_6	.
14	ν_{14}	6.4773	ν_{13}	.	ν_{13}	.	ν_{13}	.	ν_6	.
χ^2		**4.43**		.0080		.222		2.99		**26.3**

[a] Coefficients in **bold** are significant at the 5% significance level

Table 3. Two Alternative Models of Person 1's Shadow Values

	Model Number 5			Model Number 9		
Activity	Scarcity Value	Parameter Estimate[a]	Student's-t Statistic	Scarcity Value	Parameter Estimate	Student's-t Statistic
1	v_1	**-1.3175**	-3.72	v_{11}	**-1.0432**	-14.85
2	v_2	**-5.3175**	-8.89	v_2	**-5.0432**	-7.41
3	v_1	**-1.3175**	-3.72	v_{11}	**-1.0432**	-14.85
4	v_6	5.643	1.81	v_6	**8.3092**	12.04
5	v_5	6.8248	1.91	v_6	**8.3092**	12.04
6	v_6	5.643	1.81	v_6	**8.3092**	12.04
7	v_6	5.643	1.81	v_6	**8.3092**	12.04
8	v_W	**19.4124**	10.59	v_W	**20.7842**	27.26
9	v_6	5.643	1.81	v_6	**8.3092**	12.04
10	v_6	5.643	1.81	v_6	**8.3092**	12.04
11	v_{11}	-1.6194	-0.74	v_{11}	**-1.0432**	-14.85
12	v_{12}	-3.8576	-1.20	v_{11}	**-1.0432**	-14.85
13	v_{13}	2.6424	0.82	v_{13}	**5.4263**	7.62
14	v_{14}	4.0949	1.88	v_{13}	**5.4263**	7.62
Critical $t_{.05}$ (2-tailed)		2.57			2.31	
Mean log-L		-0.06783			-0.34121	

[a] Coefficients in **bold** are significant at the 5% significance level

Table 4. Person 1's Scarcity Values and Marginal Values of Activities

	Model Number 5		Model Number 9	
Activity	Scarcity Value ($/hr)	Marginal Value ($/hr)	Scarcity Value ($/hr)	Marginal Value ($/hr)
1	-1.32	0.68	-1.04	0.96
2	-5.32	0.68	-5.04	0.96
3	-1.32	0.68	-1.04	0.96
4	5.64	5.64	8.31	8.31
5	6.82	6.82	8.31	8.31
6	5.64	5.74	8.31	8.41
7	5.64	5.64	8.31	8.31
8	19.41	3.41	20.78	4.78
9	5.64	6.64	8.31	9.31
10	5.64	6.36	8.31	9.02
11	-1.62	4.09	-1.04	4.67
12	-3.86	6.14	-1.04	8.96
13	2.64	6.14	5.43	8.93
14	4.09	4.09	5.43	5.43

Table 5. Sequential Testing for the Shadow Values of Person 2

Model	1		2		3		4		5	
Activity	Scarcity Value	Estimate[a]	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate
1	ν_1	15.8718	ν_1	15.8751	ν_7	.	ν_7	.	ν_7	.
2	ν_2	12.6234	ν_2	12.6263	ν_2	**12.6221**	ν_2	12.6119	ν_2	**12.2089**
3	ν_3	6.4784	ν_3	**6.4798**	ν_3	**6.4827**	ν_3	**6.4774**	ν_3	**6.487**
4	ν_w	-2.8107	ν_w	-2.8091	ν_w	**-2.802**	ν_w	-2.8088	ν_w	**-2.7858**
5	ν_w	.	ν_w	.	ν_w	.	ν_w	.	ν_w	.
6	ν_6	13.7119	ν_6	**13.7158**	ν_6	**13.7216**	ν_6	**14.1437**	ν_6	**14.1634**
7	ν_7	15.7758	ν_7	**15.629**	ν_7	**15.678**	ν_7	**15.6727**	ν_7	**15.686**
8	ν_8	11.8785	ν_8	**11.8857**	ν_8	**11.8906**	ν_8	**11.8725**	ν_2	.
9	ν_9	4.3785	ν_9	4.3857	ν_9	4.3906	ν_9	4.3725	ν_9	4.3908
10	ν_{10}	14.6398	ν_{10}	**14.6411**	ν_{10}	**14.645**	ν_6	.	ν_6	.
11	ν_{11}	15.4731	ν_7	.	ν_7	.	ν_7	.	ν_7	.
12	ν_{12}	16.8065	ν_{12}	**16.8076**	ν_{12}	**16.8114**	ν_{12}	**16.8117**	ν_{12}	**16.8352**
χ^2		-1.1E-12		0.0072		0.0034		0.0730		0.0410

Model	6		7		8		9		10	
Activity	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate	Scarcity Value	Estimate
1	ν_7	.	ν_7	.	ν_7	.	ν_7	.	ν_7	.
2	ν_2	**12.2068**	ν_2	**13.2577**	ν_2	**13.5014**	ν_7	.	ν_7	.
3	ν_3	**6.4846**	ν_3	**6.4714**	ν_3	**6.4206**	ν_3	**6.4471**	ν_7	
4	ν_w	**-2.7951**	ν_w	**-2.8418**	ν_w	**-2.6952**	ν_w	**-2.6775**	ν_w	**-3.2325**
5	ν_w	.	ν_w	.	ν_w	.	ν_w	.	ν_w	.
6	ν_6	**14.1622**	ν_2	.	ν_2	.	ν_7	.	ν_7	.
7	ν_7	**16.0214**	ν_7	**15.9984**	ν_7	**16.2617**	ν_7	**14.9357**	ν_7	**11.9758**
8	ν_2	.	ν_2	.	ν_2	.	ν_7	.	ν_7	.
9	ν_9	4.3773	ν_9	4.3282	ν_3	.	ν_3	.	ν_7	.
10	ν_6	.	ν_2	.	ν_2	.	ν_7	.	ν_7	.
11	ν_7	.	ν_7	.	ν_7	.	ν_7	.	ν_7	.
12	ν_7	.	ν_7	.	ν_7	.	ν_7	.	ν_7	.
χ^2		0.1450		0.5986		0.6473		2.1151		**32.099**

[a] Coefficients in **bold** are significant at the 5% significance level

Table 6. Two Alternative Models of Person 2's Shadow Values

	Model Number 3			Model Number 9		
Activity	Scarcity Value	Parameter Estimate[a]	Student's-t Statistic	Scarcity Value	Parameter Estimate	Student's-t Statistic
1	v_1	**15.678**	8.00	v_7	**14.9357**	10.475
2	v_2	**12.6221**	4.29	v_7	**14.9357**	10.475
3	v_3	**6.4827**	8.02	v_3	**6.4471**	8.283
4	v_W	**-2.802**	-3.31	v_W	**-2.6775**	-3.307
5	v_W	**-2.802**	-3.31	v_W	**-2.6775**	-3.307
6	v_6	**13.7216**	5.40	v_7	**14.9357**	10.475
7	v_7	**15.678**	8.00	v_7	**14.9357**	10.475
8	v_8	**11.8906**	4.63	v_7	**14.9357**	10.475
9	v_9	4.3906	1.71	v_3	**6.4471**	8.283
10	v_9	**14.645**	5.23	v_7	**14.9357**	10.475
11	v_{11}	**15.678**	8.00	v_7	**14.9357**	10.475
12	v_{12}	**15.678**	8.00	v_7	**14.9357**	10.475
Critical $t_{.05}$ (2-tailed)			3.18			2.31
Mean log-L			-8.17729			-8.32856

[a] Coefficient estimates in **bold** are significant at the 5% significance level

Table 7. Person 2's Scarcity Values and Marginal Values of Activities

	Model Number 5		Model Number 9	
Activity	Scarcity Value ($/hr)	Marginal Value ($/hr)	Scarcity Value ($/hr)	Marginal Value ($/hr)
1	15.68	16.48	14.94	15.74
2	12.62	14.29	14.94	16.60
3	6.48	7.15	6.45	7.11
4	-2.80	-20.55	-2.68	-20.43
5	-2.80	7.20	-2.68	7.32
6	13.72	14.39	14.94	15.60
7	15.68	16.08	14.94	15.34
8	11.89	14.39	14.94	17.44
9	4.39	14.39	6.45	16.45
10	14.64	17.98	14.94	18.27
11	15.68	18.18	14.94	17.44
12	16.81	17.98	14.94	16.10

Land Use Planning and Farm Land Protection in the Bluegrass

Ronald A. Fleming
University of Kentucky
Department of Agricultural Economics

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Acknowledgements: Much thanks goes to Dr. David Debertin for his thoughts and careful editing.

Abstract: Like other places, Lexington and Fayette county, Kentucky, located in the heart of a unique thoroughbred horse producing region called the Bluegrass, is embroiled in debate concerning land use planning and farm land protection. At issue is a desire for the city of Lexington to grow while preserving a historically important agricultural industry that is also aesthetically pleasing. Past land use planning/farm land protection policies, while innovative for their time, have not stopped development of rural, horse producing areas. As a consequence other policies are being proposed. This paper discusses how the unique characteristics of the Bluegrass Region (specifically Fayette county) have acted to shape local land use policy that is vastly different from other regions of the country.

Land use issues, especially at the "ag-urban fringe" have long been a concern across the US. No US region is immune. In the west, growth of Los Angeles, Portland, Denver and other cities has resulted in the conversion of large tracts of agricultural land to other uses. The same is true in the plains, mid-west, south, and eastern states. In fact, because the largest share of the US population is located in the eastern third of the country, it might be argued that agricultural land protection and green space preservation is reaching a critical point here. And this is evidenced by the development of federal, state, and local plans to address this issue.

In the midst of the controversy is Lexington, Kentucky, the principal city in a 11 county area known as the inner Bluegrass Region. While the issues here are like those elsewhere (i.e., urban growth is taking up large tracts of otherwise productive agricultural lands), the nature of this region's people and agriculture are very different. As a consequence, land use policies that work well in other places may not be as effective here. This paper uses Lexington and the inner Bluegrass region as a case study. We will explore the history of land use preservation here, consider current issues, and discuss the likely success of policies currently in place.

Historical Background

Lexington, Kentucky instituted the nation's first Urban Service Area (Daniels and Bowers, 1997). In 1958 the city of Lexington imposed a geographic boundary to define the area beyond which the city would not provide urban services such as water, sewer, roads, and schools. Designated the Urban Service Area, this growth management policy for many years restricted the city limits of Lexington to approximately 30% of the total area of Fayette County (ES630, 2000). The primary purpose of the Urban Service Area was to protect the area's signature horse farms. Unfortunately, however, the Urban Services Area did not restrict residential development as intended.

While it is now impossible to prove, it is the opinion of some close to early development of local land use planning that the true intent of the Urban Services Area was to restrict access to city sewers. The idea was that restricting city sewer access would constrain Lexington's growth into horse producing (rural) areas and encourage compact development within the city. This strategy was basically misguided. Lexington continued to grow, but this was attributed to residential growth in Fayette County and not to growth in the city of Lexington. For many years residential subdivisions that relied on septic systems were considered part of the county, hence were technically not part of the city although adjacent to the city. So while Lexington proper was slowly growing, the irony is that adjacent county residential subdivisions were expanding rapidly to the south over traditional agricultural areas and some horse farms.

Clearly, establishing an urban service area did not stop growth outside of Lexington. Furthermore, the realities of rapid growth (124% in the county since 1950) and the politics of development have resulted in "service area creep" far beyond the boundary set in 1958. In 1973 Lexington and Fayette County were the first city and county governments in the nation to merge into a single unit. The resulting Lexington-Fayette Urban County Government (LFUCG) quickly developed a comprehensive plan to guide development by reestablishing the urban service area near its current extent and by specifying 10 acre minimum tract (lot) sizes outside of the urban service area. A minimum tract (or lot) size is a legally defined minimum area (in this context, acres) on which a residence may be constructed.

However, the greatest contribution of this new form of government may have been the elimination of what was becoming a major environmental concern and the changing of a mind set that Lexington area growth could be controlled simply by not extending sewer lines. By 1970, Lexington's land use policies had resulted in serious health problems arising from old septic systems on lots that were too small to provide an adequate drainage field. Hence the switch to an Urban County Government (UCG) was not without cost. Specifically, the LFUCG had to extend sewer lines into all neighborhoods where sewer service had not originally been offered. The bottom line is that this cost taxpayers more than it would have had proper land use planning been conducted in the first place.

As already discussed, the problem of converting horse farms into residential development is not new. Yet it is important to understand that the geographic shape (the pattern of development) of Lexington has evolved to some degree based on where horse farms were located irrespective of earlier land use planning. Specifically, south Lexington grew more rapidly than north Lexington because horse farms were more concentrated on the north side of town. The south side of town was predominately farmland (beef & tobacco land) that could be more easily developed. Today, nearly all of the land available for development in southern Fayette County has been converted, hence there is increased pressure to develop northern Fayette County, the location of many horse farms.

In April 1999, the Urban County Planning Commission considered a plan by which a tax increase would be used to pay farmers to give up their right to develop land for residential or commercial purposes. Such a program is called a Purchase of Development Rights or PDR program. The plan presented by the commission, while open to all agricultural producers, was designed to favor horse farms in general and one horse farm in particular (Calumet; LHL, 1999). Specifically, the proposal favors 1) larger horse farms, 2) horse farms with heavy improvements like specialized barns and rail fences, 3) farms with high sales, and 4) horse farms located away from the urban services area. It was noted, however, that exceptions to point 4 would be made in the case of a farm with "overwhelming importance as a community icon" (like Calumet).

To fund the program three tax increases were proposed, all of which would have to be approved by a majority vote of county residents: 1) increase property tax up to $50 per $100,000 assessed value; 2) increase the occupational (or payroll) tax of Fayette county residents by 0.125%; or 3) increase the hotel/motel room tax by 1 percent. The first two proposals directly impact Fayette county residents. An increase in the property tax rate would increase the tax on an $80,000 home by $40 a year while an increase in the payroll tax would cost a person earning the counties median annual income ($44,000) $55 a year (LHL, 1999). It is estimated that $100 million over 20 years would be required to protect 50,000 acres (27%) of "prime horse land." It is also estimated that the first two proposals would generate between 6 and 7 million per year while the third proposal would generate only $900,000. To date, no ballot measure has been presented to the public for a vote.

Farmland Versus Open Space Preservation

Given the stated purpose of the proposed PDR program being to protect agriculture in Fayette County, it would appear that the LFUCG is highly interested in farmland protection. In fact, press releases by the urban county government and others often refer to the protection of farms as being a noble and just cause. Yet is it really agriculture and farmland that the urban county government wants to protect or is it open space in general? It is argued that Lexington homeowners do not care if horses or any other form of agriculture occupies space; they only care for open space and scenic vistas. If it is agriculture that the urban county government is trying to protect, then it is equally true that it is a particular type of agriculture that is being saved. Specifically, it is the large horse farm with specialized barns and miles of plank fencing that are target for protection. These are the farms that people see as they drive or fly into town, the farms that epitomize the horse industry of the Bluegrass. Yet is this truly farm land protection or is this protection of a unique industry that happens to be agricultural in nature, but is vastly different from the agriculture that is typically in mind when one talks of farmland preservation?

In fact, while not yet a major concern in Fayette County, there is much vocal opposition to what might be classified as industrialized hog or chicken operations across Kentucky. Would the proponents of farmland preservation in Fayette County be willing to protect agricultural land used for confined animal production? While it is not possible to give a definitive answer, the preponderance of evidence from various media sources would suggest that protection of such firms is not the intent of land use planners, nor would such protections be well received by a majority of the public. Therefore, it cannot be argued that the proposed farm land protection policies are meant to apply to all agricultural producers; just those that meet the public's perception of "Bluegrass agriculture." The reality is that confined animal feeding operations (CAFOs), although legitimate agricultural producers, need not apply.

Daniels and Bowers in their book state that the protection of nonfarm open space often involves smaller properties, a wide variety of landowners, and issues of parks, trails, greenways, public recreation, and wildlife habitat. Farmland protection, on the other hand, may be though of as open-space protection without public access to the property. On this basis alone, it would appear that the urban county government is engaged in farmland protection. In its current form, the proposed PDR program is voluntary and allows owners to retain title to the land. The land does not become a public park. Yet there are other elements of a farmland protection program. Specifically, the land area protected should be sufficiently large to allow individual farms to attain the size they need to be profitable and there should be enough of these farms to support secondary service and marketing firms. Furthermore, the program should be designed to allow farmers to expand or change their operation to remain competitive and it should help farmers avoid conflicts with non-farm neighbors (over dust, chemical use, noise, odors, etc.). How does the Fayette County plan measure up to these ideals?

In 1997, Fayette County was number one in the state in cash receipts from farm marketing ($365,971,000; KAS, 1999). While exact figures are not reported, Kentucky Agricultural Statistics (KAS) does report that the sale of horses and stud fees accounted for most of the livestock cash receipts for Fayette county (KAS, 1999). Furthermore, livestock receipts represent 94% of all agricultural cash receipts for the county. Table 1 reports relevant agricultural information for 1992 and 1997. In 1992 there were 836 farms in Fayette county of which 72 (8.67%) produced corn, 19 (2.3%) produced wheat, 24 (2.9%) produced soybeans, 484 (58%) produced tobacco, 297 (35.5%) produced cattle, and 12 (1.4%) produced hogs. Interestingly, while 94% of all agricultural cash receipts in 1997 were derived from livestock production in 1997, only 37% of Fayette county farms were engaged in cow, dairy, and (or) swine production. Clearly, horse production is driving livestock production and sales in Fayette County. By 1997 these numbers had fallen to 745 total farmers (an 11% reduction) of which 51 produced corn (a

29% reduction), 23 produced wheat (a 21% increase), 32 produced soybeans (a 33% increase), 340 produce tobacco (a 30% reduction), 269 produce cattle (a 9% reduction), and only 3 farmers produce pigs (a 75% reduction).

Because farm numbers are falling, it would be easy to say that farmland preservation programs are not working. However, without further information, it is not possible to separate those farms that were developed from those that went out of business for economic reasons and were purchased by other farmers. For example, low market prices may be the primary reason for the decline in overall farm numbers and for the decline in the number of farmers producing corn, tobacco, cattle, and hogs (Table 1). Yet, Table 1 also shows that total acres in farms also decreased between 1992 and 1997 by 10,231 acres (7%). Given an average farm size of 182 acres (USDA-NASS, 1999) this represents 56 farms.

It would be hard to deny that, at least, the non-horse agriculture in Fayette county is under some development pressure. Between 1992 and 1995 it was farms less than 180 acres and greater than 1,000 acres that shut down (Table 1). However, it is equally true that the story may be different for horse production. From 1963 to 1995 acreage in horse farms has increased 55% from 34,197 to 53,128 acres (LHL, 1999). Furthermore, there were 170 more horse farms in 1997 (375) than there were in 1992 (205), an 83% increase in the number of farms (USDA-NASS, 1999).

For commodity groups other than horse production, the proposed PDR program is likely to be too little, too late. Despite prior planning, development continued in rural areas, fragmenting the landscape and isolating "islands" of farmland that are surrounded by developed areas. The number of commercial farm operations has dwindled having been replaced with residential mini-farms and estates. This is a direct consequence of regulation restricting tract sizes to 10 acres or less outside the urban services area. Instead of limiting growth in rural areas, this policy acted to accelerate it by encouraging conversion of large tracts of agricultural land into 10-acre rural estates (what is referred to as estate sprawl later). With fewer commercial farmers, many farm service and marketing firms have chosen to close their businesses. Those that survived adopted to serve the needs of what are now largely hobby farmers.

Hence, in summary, earlier policies that established the urban service area and minimum tract size restrictions acted to accelerate development in rural areas to the determent of traditional commercial agriculture. Open space with attached scenic amenities was preserved, although not in a form that can be used by the public (i.e., the open space is still privately owned). A PDR program can help those traditional commercial farms that remain, but these farms will struggle to remain economically viable as they have to increasing deal with the concerns of their non-rural neighbors and have to travel farther for services.

While a comprehensive plan to protect traditional commercial agriculture in Fayette county may be futile, there is evidence that prior land use policies did aid the horse industry. The impact of the proposed PDR program, however, is suspect. A popular perception is that development has consumed the horse farms of Fayette County. Examples include Hamburg Place, Beaumont Farm, and Coldstream Farm, which have been partly to fully developed. However, large parts of these farms were inside of the 1958 urban services area, hence slated for future development (LHL, 1999). Yet, since 1963, there has been a steady increase in the acreage owned by horse farms from 34,197 acres in 1963 to 53,128 acres in 1995 (LHL, 1999). In fact, the data suggests that horse farms have concentrated in the north and northwestern parts of the county. Expansion and concentration took place via conversion of other types of agriculture into horse farms. In this way, established horse farms "moved themselves" away from the urban services area and new farms were established. In summary, it appears that Fayette county's land use policies did act to protect the county's horse farms and had, in fact, the desired qualities of concentrating the industry in an area large enough that individual firms could continue to operate profitably without being hassled by urban residents and expand if desired. And the industry is large enough to support necessary, local service and marketing firms.

Yet, the following questions remain. First, did land use planning really aid the horse industry or was the horse industry simply rich enough to make the adjustments (buy land etc.) they needed to survive? Second, will adding a PDR program and increasing minimum tract size from 10 to 40 acres really benefit the horse industry? Again, we have established that Fayette county's land use policies have not aided traditional crop and livestock commercial agriculture. These farms and their supporting service continue to disappear. However, we have also established that what has happened is that traditional crop and livestock farms have been converted to horse farms. Hence, Fayette county's land use policies have aided agriculture, albeit, a special form of agriculture, horse farming. But then, this is the stated intent of past and recent land use planning.

However, it is my contention that the trends in growth and concentration of the county's horse industry would have occurred with or without land use planning. The missing element often not discussed is the great wealth of the owners of many of, at least, the larger horse farms. Ownership of horse farms in Fayette county is largely by very wealthy peoples some of which have permanent residence elsewhere. Hence, land use planning here is not about the protection of farmers struggling to keep their land and their livelihood as it is in many other places of the country. Land use planning is about maintaining an industry and, arguably, an industry associated with the very wealthy of society. For this reason, many of the wealthy horse farmers such as William T. Young of Overbrook Farm and Sheik Hamdan bin Rashid of Shadwell Farm are not interested in the PDR program (LHL, 1999). They grew their operations by acquisition, hence are largely unaffected by and ambivalent toward land use planning. So who is it that will benefit from the proposed PDR program? It is the smaller, more financially limited horse farms who must compete against the larger horse farms and who feel most threatened by development. They are stuck where they are and cannot move by acquisition. It is also the more profitable horse farms that want to insure enough acreage so that the industry can continue to grow. If farmers use the PDR program to protect their land from development, then the horse industry knows that land will be available for future expansion.

What is urban sprawl?

If the goal of land use planning, and the PDR program in particular, is to protect the future of the horse industry by insuring enough acreage so that the industry can grow, then, from an agricultural policy point of view, it must be argued that such planning is a worthwhile goal and that it has been relatively successful. However, it is also fact that the planning policies of Fayette County have yet to stop residential development beyond the urban service area. The horse farms of Fayette county, although more prevalent and concentrated away from the urban service area, are still infiltrated by pockets of residential development. This, I believe, is a direct consequence of minimum tract (or lot) restrictions. Furthermore, I am not convinced that a recently passed proposal to increase the minimum tract size on which a residence may be built from 10 to 40 acres will benefit the horse industry.

Urban sprawl is generally viewed as tract housing and strip type shopping centers on what was once agricultural ground or, at least, open space located at an increasing distance from the urban center. I do not dispute this definition. Furthermore, it is well known that this type of development has consumed a great deal of prime agricultural areas, has lead to increase reliance on cars which has lead to traffic, noise, and pollution, has lead to increased urban flooding (ES630, 2000), and has lead to increased service loads on urban infrastructures such as sewer systems, water systems, and school districts (Daniels and Bowers, 1997). While this type of urban sprawl is evident in Fayette County, I argue that the creation of landed estates via minimum tract sizes is, from an agricultural point of view, just as troubling a form of sprawl as is tract housing. I call this "estate sprawl."

Again, following Daniels and Bowers (1997), the purpose of farmland preservation is to protect a sufficiently large area of agricultural land that sustains a sufficiently large farm population to insulate itself from urban complaints (noise, odor, dust, equipment on road ways, etc.) and support surrounding service firms currently and in the future. The use of minimum tract sizes is a land use strategy designed for this purpose. For effective agricultural land protection, Daniels and Bowers, suggest 80-acre minimums. The unfortunate reality is, however, that most minimum tract sizes are substantially less than 80 acres, usually 10 to 30 acres. In more wealthy regions, urban people seeking a rural setting are able to buy up these tracts. In this way agricultural areas are fragmented by pockets (or clusters) of residential holdings. In my opinion, these are nothing less than landed estates for the wealthy of society. These areas are rarely used for commercial agricultural purposes and only seldom used to support agriculture of any kind. Over time this intrusion into agricultural areas becomes increasing disruptive to commercial agriculture, consumes the available agricultural land base, and results in the ultimate demise of commercial agriculture. Hence, my argument that minimum tract sizes lead to urban "estate" sprawl. It may be a more esthetic form of sprawl because there is more open or green space between each residence and the homes tend to be more expensive. But this is sprawl nevertheless because what was once land supporting a viable commercial agriculture is no more. It is true that this land is more readily converted back to agricultural purposes than are track housing or strip shopping centers.

To better illustrate my problem with small (less than 80 acre) minimum tract sizes as an agricultural land preservation strategy in an affluent society, consider the following. Suppose that there are 10 families wanting to purchase land in a rural setting who have the financial means to do so. A 10-acre minimum tract size means that 100 acres of agricultural land will be consumed. Raise the minimum tract size to 40 acres and all that changes is that 300 additional acres are consumed. In an affluent society raising the minimum tract size by several 10's of acres will not pose a sufficiently large financial burden as to force such people out of the market. In fact, so few people will be turned away that it is likely that more land, not less, will be converted to residential uses (landed estates). Combine this with further lack of land use planning and the result is pocket residential areas that further fragment the countryside and make commercial farming all that more difficult. This is the case in the Fayette County. Via this process large tracts of prime agricultural land have been converted into landed estates.

In a relatively affluent area such as the Bluegrass, the political reality is that the desirable goal of preserving larger tracts of agricultural land may come up against the equally desirable goal of having development occur in a few large clusters rather than in many small clusters scattered about the countryside. In fact, if the voters of Fayette county are truly interested in farmland protection, then it would probably be better if the LFUCG abandoned it's blind reliance on minimum tract sizes as a land use policy and instead engaged in rezoning. With rezoning more land could be included into the urban services areas making it possible to develop more contiguously and compactly rather than in sparse clusters thereby reducing the total acreage converted to residential property. A reason for development of rural tracts is that city sized lots in nicer locations within the growth boundary cost nearly as much as some 10 and 40 acre tracts in the rural areas of the county. Expanding the growth boundary could reduce the price of city lots and encourage inward development and, perhaps, redevelopment of previously developed "inter-city" property. Furthermore, rezoning would allow the county to coordinate a comprehensive, area wide development plan with surrounding counties who face similar growth challenges.

Equity Issues and Other Final Thoughts to Consider

An unfortunate truth is that Fayette county farmland preservation policies, including the purchase of development rights (PDR) program, are not about preserving operational economic farms, but rather about preserving an urban "myth" about what farming in the Bluegrass was once like. The economic argument (we need farm production) is a non-starter. Residential development for a high tech firm makes far more economic sense. The problem is that, while horse farms do provide a public good, it is impractical to manage horse farms like a National Park. The PDR program will keep horse farms that provide a public goods component in private hands.

A central question surrounding farmland protection/land use policy in Fayette County is should the average citizen, especially the poor, be held financially responsible. Recall that current funding options for a PDR program (estimated to cost $100 million over 20 years to protect 50,000 acres) include an increase in property taxes, an increase in the occupational (or payroll) tax paid by Fayette county residents, or an increase in the hotel/motel room tax. Proponents of preservation (the PDR program) argue that the (unspecified and unpriced) benefits of farmland protection accrue to all Fayette county residents, hence all residents should pay for preservation. However, it cannot be denied that the program is designed to protect the larger horse farms in particular and that it is these horse farms that have the least interest in the program (LHL, 1999). It also cannot be denied that while Fayette county is populated by individual of extreme wealth, it is also populated by even more people in extreme poverty. Why should the poor of Fayette County be asked to fund a program via their taxes that the more wealthy recipients of this program marginally want? If there is reluctance on the part of Fayette county voters to agree to tax increases that subsidize wealthy rural Fayette Countians to not do something they might not have done anyway, that reluctance is well founded.

Another problem is that it will likely take much more than $100 million to preserve 50,000 acres in via tax dollars in Fayette county. Using these figures, the LFUCG intends to pay $2,000 per acre on average for acquisition of development rights. This gives rise to a number of problems. First, given that the total amount of the program is received over 20 years, only a few purchases could be made each year. The problem is that the development value of the remaining parcels would, probably, be increasing with each passing year raising the per acre cost of the program. Second, the financial incentives are such that those that own farmland in Fayette county and who never intended to develop their land would still be expected to participate in the program. In fact, it is likely that these individuals will represent the greatest demand for the program. On the other hand, owners of land on which development is likely will probably find the value of the development right worth much more than what the LFUCG is willing to pay for it. Hence, total demand for the program and actual acreage protected is likely to be limited. The end result is that the relatively more wealthy of Fayette county, who had no intention of developing, are paid by taxpayers to participate in a program with a limited impact on development.

Another fascinating issue is that, to the extent that some farmers agree to give up their development rights while others do not, the conditions are right for accelerated leapfrog development. This form of development is especially expensive as water and sewer lines pass by protected land to reach unprotected land being held for development. Are the citizens of Fayette County better off for being able to drive through green space preserved by a very expensive farmland preservation program only to run up against the type of development they thought they had prevented? I think not. And whether a 40-acre estate is prettier than a 10-acre estate is truly a value-laden question. What is certain is that these 10 and 40-acre tracts will not, except in rare cases (i.e., horticulture), be commercial farms. Some may participate in "hobby" farming activities, but they are not likely to be viable commercial farms. For these reasons comprehensive planning and zoning is needed to concentrate residential development and protect large contiguous tracts of agricultural lands.

In conclusion, while cast as farmland preservation, Fayette county's land use policies and the purchase of development rights program in particular is not about farmland protection in the strictest sense. This program may allow some current middle and small farms to protect their land from development, but the program is not sufficiently large nor sufficiently ahead of current development trends as to aid continued survival of these farms. The proposed program is simply not sufficient to support (or help grow) a viable commercial agriculture. Some farms will survive, but only those willing to continue operation in the face of problems associated with continued development. The current policy restricting land purchases to 40 acre tracts is also not a farmland protection policy. This policy will result in open space, albeit without public access, but it will not result in protected farmland. Instead, these tracts will be converted to landed estates (estate sprawl) by the wealthy.

At it's core, land use planning in Fayette county is not about farmland protection, but about protecting the horse industry, an industry associated with the very wealthy of society. And what fails to be recognized is that this industry is largely ambivalent toward land use planning as they have the financial resources to acquire land and expand as needed. As such, it is difficult to argue that these are programs that should be financed by tax payers, especially the financially disadvantaged peoples of the county. Instead, the LFUCG should engage in wholesale rezoning where land already consumed by leap-frog rural residents is allowed to develop and other areas are restricted to agricultural practices. In this way sufficient contiguous land might be set aside to provide for a viable agricultural community. Unfortunately, given past and current land use policies and the resulting loss of the agricultural community and its support services, it may be too late for all but the more wealthy Fayette county farms.

Epilogue

In late May of 2000, the Mayor of Lexington proposed that 15 million dollars of phase one master (tobacco) settlement funds be used support a pilot PDR program in Fayette county. An additional 15 million dollars of matching funds would be raised through bonds issued by the city. The governor of the state favors use of master settlement funds for land use preservation programs, hence supports Fayette county's proposal for funds. In January 2001 Fayette County's request for 15 million from master settlement funds was approved by the state board overseeing these funds despite concerns raised that funds meant to aid tobacco producers were being used to fund a program perceived to benefit wealthy Lexington horse farmers. Approval was granted because of the strong support shown by the governor and other key legislators for this program.

While, perhaps, a set in the right direction given that those who are willing to pay for land use preservation are able to freely do so through the purchase of a bond, this proposal is still limited. First, the original goal of protecting 50,000 acres of "prime" horse acreage was estimated to cost $100 million over 20 years. The current proposal covers only a third of the anticipated cost, albeit, as a one time fund. The LFUCG has not specified how they intend to raise the other $50 million, nor have they specified the time frame for raising such support. Second, local "experts" close to Fayette county realestate values argue that the $2,000 per acre that the LFUCG plans to pay for development rights is simply not realistic. This linked with lower funding level clearly implies that much less land than anticipated will be protected by the program. This further erodes the capacity of the program to protect acreage in a sufficiently large block as to avoid pressures associated rural residential development. Finally, the proposed PDR programs do not remedy past planning errors (may in fact reinforce these errors) and allows the LFUCG to continue to ignore the need for comprehensive land use planning.

References

Daniels, Tom, and Deborah Bowers. Holding Our Ground: Protecting America's Farms and Farmland. Island Press. Washington, D.C. 1997.

Environmental Systems 630 (ES630). "Case Study: Stormwater Management Systems in Lexington-Fayette County's Expansion Area 2." Student paper prepared and presented by Ann Daugherty, Jan Jennemann, and Scott Phelps. Dr. Tomas Greider Class Coordinator, University of Kentucky, Spring Semester, 2000.

Kentucky Agricultural Statistics Service (KASS). "Kentucky Agriculture Statistics 1997-1998." Annual Report issued cooperatively by the National Agricultural Statistics Service and the Kentucky Department of Agriculture. Louisville, Kentucky. 1999.

Lexington Herald Leader (LHL). "Common Ground: Deciding how the Bluegrass should grow." A series of special reports. Starting Sunday, April 4, 1999.

United State Department of Agriculture (USDA), National Agricultural Statistics Service (NASS). "1992 Census of Agriculture: State and County Highlights Data." www.nass.usda.gov/census/census92/datafile/Kyc034.txt. 1994.

United State Department of Agriculture (USDA), National Agricultural Statistics Service (NASS). "1997 Census of Agriculture: U.S. State - Level Data." Table 1. County Summary Highlights: 1997. Page 5. www.nass.usda.gov/census/census97/volume1/Ky-17/Ky2_01.pdf. 1999.

Table 1. 1992 and 1997 Farms, Production, and Acreage Information for Fayette County Kentucky.

	Farms		% Change[c]	Acres or Head		% Change
	1992[a]	1997[b]		1992[a]	1997[b]	
Total Farms	836	745	-10.9	147,154	136,923	-7.0
Farms: 1 to 9 Ac.	88	70	-20.5			
10 to 49	287	248	-13.6			
50 to 179	241	215	-10.8			
180 to 499	148	148	0.0			
500 to 999	45	45	0.0			
≥ 1,000 Ac.	27	19	-29.6			
Farms Growing:						
Corn	72	51	-29.2	4,024	3,594	-10.7
Wheat	19	23	21.1	803	762	-5.1
Soybeans	24	32	33.3	1,795	2,283	27.2
Tobacco	484	340	-29.8	6,285	4,863	-22.6
Cattle	297	269	-9.4	22,320	24,855	11.4
Hogs	12	3	-75.0	**	**	**

a. USDA-NASS, 1999
b. USDA-NASS, 1994
c. Percentage Change (% Change) is calculated as ((1997 - 1992)/1992)*100.
** Data not reported by NASS in 1997 due to confidentiality concerns.

Addressing Aggregation Bias in Zonal Recreation Models

Klaus Moeltner
Department of Applied Economics and Statistics
University of Nevada, Reno

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

The author thanks Jeffrey E. Englin, W. Douglass Shaw, and J. Scott Shonkwiler for insightful comments. Address correspondence to Klaus Moeltner, Department of Applied Economics and Statistics, University of Nevada / Reno, Mail Stop 204 / Reno, NV 89557-0105, Phone: (775) 784-6701 (work), (775) 827-8949 (home), Fax: (775) 784-1342, e-mail: moeltner@unr.edu

Abstract: Models of recreation demand are often based on zonal data. Results from such models are susceptible to aggregation bias. We propose a zonal model of recreation that captures some of the underlying heterogeneity of individual visitors by incorporating distributional information on per-capita income from Census sources into the aggregate demand function. This adjustment eliminates the unrealistic constraint of constant income across zonal residents, and thus reduces the risk of aggregation bias in estimated parameters. In addition, the corrected aggregate specification reinstates the applicability of generalized maximum likelihood methods, and increases model efficiency.

Key words: Aggregation bias; Count data models; Generalized maximum likelihood

Introduction

Resource managers and researchers are often interested in estimating welfare to visitors generated by an entire system of recreation sites. To aid in policy evaluation, budget allocation, and cross-system comparison, these measures are often most meaningful to planners if derived at an annual or seasonal level.

In recent years, researchers have begun to combine count data specifications with utility-theoretical frameworks, such as incomplete demand systems (LaFrance and Hanemann [15]), to estimate seasonal demand for multiple recreation sites (e.g. Englin et al. [4], Shonkwiler [20], Shonkwiler and Englin [21]). Most of these studies resort to recreation permits as their main source of data. Such permit data are commonly used in the economic analysis of recreation activities that requires visitors to register with local Public Land managers at each trip occasion. These data are routinely collected by authorities, and thus constitute a convenient and inexpensive source of information to the analyst. The main drawback of using permit data, however, is that visitors' characteristics, such as income, age group, or education level, are only available at an aggregate geographical level from public sources such as national census data. Depending on the scope and objective of a given study, the relevant geographical units chosen by researchers are frequently ZIP code areas (e.g. Englin et al. [4], Hilger and Englin [12], Lutz et al. [17], Shonkwiler and Englin [21]), counties (e.g, Cicchetti et al. [3], Hellerstein [9]), or other census-type units.

As is well known in econometric estimation, the use of such aggregate data in lieu of individual-specific information will generally not yield parameter estimates that accurately reflect individual behavior. This shortcoming is commonly referred to as "aggregation bias" (e.g. Stoker [24], Zellner [26]). The aggregation problem has, to date, not found much attention in the recreation literature. Notable exceptions are Hellerstein [9] and [11], although these studies focus more on the benefits of using aggregate data in combination with robust estimation methods rather than on the nature and problems related to aggregation bias.

The aim of this paper is to examine the econometric issues associated with the use of aggregate data within the context of multi-site recreational demand and welfare estimation. It will be shown that aggregation problems can be alleviated by incorporating additional zonal information into count data specifications. This remedial technique also increases model efficiency.

The remainder of this text is structured as follows: The following section briefly discusses some relevant econometric theory associated with the aggregation problem, as well as econometric issues that arise when aggregate data is used in the context of count data models. The next segment delineates a "corrected" count data model that reduces the undesired effects of aggregation bias. The empirical part of this study then discusses data, estimation results, and welfare measures. It is followed by a sensitivity analysis of welfare results. Concluding remarks and a summary of key findings are given in the last section.

Using Aggregate Information in Count Data Models – Econometric Considerations

Econometricians have been aware for a long time that econometric models using aggregate information, such a averages or totals over sub-groups of individual observations, generally fail in systematic ways to take account of underlying heterogeneous behavioral reactions. This makes it often difficult to assign a meaningful interpretation to estimated model parameters. Stoker [24] elaborates on this notion using numerous examples from the empirical literature. He refers to the condition under which an aggregate specification unambiguously identifies an underlying micro function as "completeness", and the ability of retrieving individual-specific parameters from an aggregate model as "recoverability". In Stoker [23] he proposes the interpretation of the aggregate relationship as a reduced-form model, and the recoverability of individual- level, structural coefficients as an identification issue.

In the context of recreation demand, an unobserved micro-function could take the following general form:

$$d_{ij} = f(x_{ij};\beta), \tag{1}$$

where d_{ij} is an individual's demand (in visits) for site j, x_{ij} is a vector of site and individual characteristics, and β is a vector of parameters. The corresponding aggregate or macro function could then be derived from (1) as

$$d_{Ij}(\beta,\theta) = E_x(d_{ij}) = \int f(x_{ij};\beta)\,p(x_{ij};\theta)\,dx \ , \tag{2}$$

where the capital I in d_{Ij} indicates that the elements of x_{ij} have been aggregated over some sub-group I (e.g. a ZIP code area, or some other geographical unit). The x-subscript to the expectation sign indicates that mathematical expectations are taken over regressors x_{ij}, which follow a probability density function p with parameter vector θ. Thus, d_{Ij} can be regarded as demand by a representative agent from zone I.

The core segment of Stoker [23] identifies restrictions on (1) and $p(x;\theta)$ for completeness to hold. The key result for this study is that if $p(.)$ belongs to the family of exponential distributions, completeness is ascertained for any functional form of (1). In addition, micro-level coefficients can generally be recovered from (2) for such models if information on θ is available.

Hellerstein [11] illustrates this notion within the framework of a simulated count data model for recreation demand by assuming a joint-normal distribution for his two right hand side variables, travel cost and income. He shows how this additional information can be incorporated into the specification of an aggregate Poisson model. In a variety of simulation settings, he demonstrates that this "corrected" zonal model yields more accurate welfare estimates than generic aggregate specifications. He also finds that for many of his visitation scenarios the corrected and, in some cases, even the standard aggregate model perform better than their counterparts based on limited individual observations.

The latter result indicates that there are inherent potential benefits of using aggregate regressor information in count data models. Unlike specifications based on individual information zonal models generally do not need to be corrected for truncation and / or endogenous stratification (e.g. Hellerstein [10], Shaw [19]), since information on non-participants in aggregate form is readily available from census data at any desired geographical level. Thus, they do not require additional distributional assumptions on the dependent variable associated with such adjustments, and avoid reliance on estimators that are highly sensitive to model misspecification. Furthermore, as discussed in [9] and [11], non-limited count data models lend themselves to estimation through generalized maximum likelihood (ML) techniques, such as pseudo- and quasi-generalized pseudo ML methods (PML and QGPML, respectively - Gourieroux et al. [6] and [7]).[1] These robust estimators offer additional insurance against misspecification of the distribution of the dependent variable, and are often computationally less burdensome than standard ML methods. However, the requirement for PML and QGPML to generate consistent parameter estimates is the availability of a consistent estimate of the moments of the dependent variable. These will not be available if aggregation bias is present, which effectively renders these useful estimation techniques inapplicable for most zonal models.

From the discussion above it is clear that in order to reduce the risk of aggregation bias, and to fully exhaust the benefits in zonal models, additional information on the distribution of explanatory variables is needed. The following section shows how estimates of distribution moments for some commonly used regressors in recreation demand models can be extracted from Census data and incorporated into a count data framework.

A Corrected Aggregate Count Data Model of Recreation Demand

We apply the Incomplete Demand System framework discussed in LaFrance [14], and follow Hilger and Englin [12], Shonkwiler [20], and Shonkwiler and Englin [21] by choosing a semi-log form for the expected demand equations:[2]

$$E\left(d_{ij}\right)=\lambda_{ij}=\exp\left(\alpha_i+\beta'_{q,j}\cdot q_j+\beta_{p,ij}\cdot p_{ij}+\beta'_{p,ik}\cdot p_{ik}+\beta_{m,i}\cdot m_i\right), \qquad (3)$$

where d_{ij} is the actual demand by individual i for site j, α_i is an individual-specific intercept, q_j is a vector of trail features, p_{ij} is the price of site j to individual i measured as travel cost from i's residence to trail j, p_{ik} is a vector of prices to all other sites, m_i denotes individual income, and the β-terms are coefficients. Specifically, $\beta_{p,ij}$ is the own-price coefficient for site j, and $\beta_{p,ik}$ is a vector containing all cross-price coefficients. Imposing the cross-equation restrictions given in LaFrance [14], and the additional simplifying constraints of origin- and destination-invariant travel cost coefficients and preferences for trail features, (3) reduces to

$$E\left(d_{ij}\right)=\lambda_{ij}=\exp\left(\alpha_i+\beta'_q\cdot q_j+\beta_p\cdot p_{ij}+\beta_m\cdot m_i\right) \qquad (4)$$

Based on this specification, the IDS system yields the following expressions for expected CV and EV, assuming positive income effects (LaFrance [14]):

$$\begin{aligned} E\left(CV_i\right)&=\left(\frac{1}{\beta_m}\right)\cdot\ln\left[1+\frac{\beta_m}{\beta_p}\cdot\sum_{j=1}^{J}\left(a_{ij}\cdot\exp\left(\beta_p\cdot p^\circ_{ij}\right)-a_{ij}\cdot\exp\left(\beta_p\cdot\tilde{p}_{ij}\right)\right)\right] \\ E\left(EV_i\right)&=-\left(\frac{1}{\beta_m}\right)\cdot\ln\left[1-\frac{\beta_m}{\beta_p}\cdot\sum_{j=1}^{J}\left(a_{ij}\cdot\exp\left(\beta_p\cdot p^\circ_{ij}\right)-a_{ij}\cdot\exp\left(\beta_p\cdot\tilde{p}_{ij}\right)\right)\right], \end{aligned} \qquad (5a,b)$$

where p°_{ij} is the price of site j to individual i in the original state, and $\tilde{p}_{ij}$ is the price in the new state. The two equations constitute individual, per-season welfare measures resulting from a price change for one or more destinations within the recreation system. To capture the entire area under the compensated demand curve associated with either CV or EV, one can set $\tilde{p}_{ij}$ to infinity. In that case the last terms in (5a,b) vanish.

In the spirit of Stoker's terminology presented above, (4) constitutes the unobserved micro-function for this model. Only trail features (i.e. the elements of q_j) are known, while aggregate values need to be substituted for α_i, p_{ij}, and m_i. As discussed below in more detail, aggregate information was collected for 134 ZIP-codes for this study. In order to keep the number of parameters to be estimated within reasonable limits, these ZIP code areas are further grouped into 35 population zones (PZ) for some parts of this analysis.

Since the micro-function is non-linear, the recoverability of the coefficients in (4) hinges on assumptions regarding the distribution of the three variables in question within a given ZIP code. Since no such information can be gleaned from public data sources for α_i, we simply assume that all residents of a given PZ share the same preferences as reflected by the intercept term. In other words, we stipulate that $\alpha_i = \alpha_Z, \forall i \in Z, Z = 1 \cdots 35$, where Z denotes a given population zone.

The computation of an aggregate value for p_{ij} requires a closer look at the exact definition of this variable. To facilitate the comparison of results generated by this model with those found in other studies, we follow the conventional approach of defining travel cost as a combination of mile cost and time cost, using a third of the hourly wage rate to capture the opportunity cost of time. Specifically,

$$
\begin{aligned}
p_{ij} &= ml_{ij} + f_{ij} + \gamma \cdot m_i \cdot t_{ij}\,, \quad \text{with} \\
ml_{ij} &= \frac{(2 \cdot \delta_{ij} \cdot \$0.25)}{g_{ij}} \\
\gamma &= \frac{0.33}{2080} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(6a-d)} \\
t_{ij} &= \left(\frac{\delta_{hwy,ij}}{60}\right) + \left(\frac{\delta_{acc,ij}}{30}\right),
\end{aligned}
$$

where ml_{ij} are mile costs based on a vehicle operating expense of \$0.25/mile,[3] δ_{ij} is the distance, in miles, from i's residence to site j, and g_{ij} is the number of passengers traveling in the same vehicle. Time cost, t_{ij}, is divided into a highway and an access-road component, where an access road is defined as the trip segment between highway exit and final destination. As reflected in (6d), we allow for twice as high a speed level on highways than on access roads. For some origins, visitors also incur ferry fees on their way to the recreation system. For these ZIP areas, an appropriate cost term, f_{ij}, was added to (6a). Since information on the exact location of a visitor's residence is not available, we make the simplifying assumption that all residents of a given ZIP area face the same distance in miles to a specific site within the recreation system of interest, i.e. $\delta_{ij} = \delta_{Ij}, \forall i \in I, I = 1 \cdots 134$. Also, we substitute the average group size for a given ZIP as captured by permit data for g_{ij}, i.e. $g_{ij} = \bar{g}_{Ij}, \forall i \in I, I = 1 \cdots 134$.

To this point, all components of the micro-function (4) are assumed to be equal for all residents of a specific ZIP area. In a simple aggregate model, one would extent this notion also to personal income, m_i, and estimate the following equation for the expected demand of a representative individual from zone I for destination j:

$$\tilde{\lambda}_{Ij} = \exp\left(\tilde{\alpha}_z + \tilde{\beta}'_q \cdot q_j + \tilde{\beta}_p \cdot \overline{p}_{Ij} + \tilde{\beta}_m \cdot \overline{m}_I\right), \tag{7}$$

where $\overline{m}_I$ is the mean per-capita income for ZIP I, and $\overline{p}_{Ij}$ is given by (6a) with implementation of distance and group size constraints, and m_i replaced by $\overline{m}_I$. Since the assumption of zero variance for income is unrealistic, such a "naïve" model will most likely be flawed by aggregation bias. The resulting parameter values and predictions for site demand will not be consistent estimates for the ones given in (4). This is reflected by the "~" superscripts for coefficients and expected demand in (7).

However, information on the distribution of m_i for a given ZIP area can be extracted from common Census data sources. This additional information allows for the specification of an improved aggregate model. Based on a graphical inspection of income histograms, we stipulate that, for each ZIP area, income is distributed normally within two population groups: white ("w") and non-white ("nw"). Since the normal distribution belongs to the exponential family, this specification satisfies Stoker's [23] completeness condition mentioned above. Income, through the hourly wage rate, also enters the definition of travel cost (see 6a), which implies that the two regressors are distributed joint-normally for each race group, i.e.

$$\begin{bmatrix} p_{ij,r} \\ m_{i,r} \end{bmatrix} \sim mvn\left(\mu_{Ij,r}, \Omega_{Ij,r}\right), \qquad \text{where}$$

$$\mu_{Ij,r} = \begin{bmatrix} nt_{Ij} + \gamma \cdot t_{Ij} \cdot \overline{m}_{I,r} \\ \overline{m}_{I,r} \end{bmatrix}, \quad \Omega_{Ij,r} = \begin{bmatrix} \left(\gamma \cdot t_{Ij}\right)^2 \cdot \sigma^2_{I,r} & \gamma \cdot t_{Ij} \cdot \sigma^2_{I,r} \\ \gamma \cdot t_{Ij} \cdot \sigma^2_{I,r} & \sigma^2_{I,r} \end{bmatrix} \tag{8}$$

$$nt_{Ij} = ml_{Ij} + f_{Ij}$$

$$r = w, nw$$

The notation is as follows: mvn denotes the multivariate normal distribution, nt_{Ij} captures non-time costs (mile and ferry expenses), $\overline{m}_{I,r}$ and $\sigma^2_{I,r}$ are the mean and variance of income for ZIP I and ethnic group r, and the remaining terms are defined in (6c,d). As discussed in Hellerstein [11] the exponents of travel cost and income for each race group will be jointly log-normally distributed with the following mean and variance:

$$E\left(\exp\begin{bmatrix} p_{ij,r} \\ m_{i,r} \end{bmatrix}\right) = \exp\left(\mu_{Ij,r} + \frac{1}{2} \cdot \Omega_{Ij,r}\right),$$

$$V\left(\exp\begin{bmatrix} p_{ij,r} \\ m_{i,r} \end{bmatrix}\right) = \exp\left(2 \cdot \mu_{Ij,r} + 2 \cdot \Omega_{Ij,r}\right) - \exp\left(2 \cdot \mu_{Ij,r} + \Omega_{Ij,r}\right) \tag{9}$$

$$r = w, nw$$

The census also reveals the population share for each ethnic segment. Using this information in combination with (8) and (9), we can define a corrected macro-function for expected individual site demand as

$$
\begin{aligned}
\lambda_{Ij} &= E_x\left(\lambda_{ij}\right) = pw_I \cdot \lambda_{Ij,w} + \left(1 - pw_I\right) \cdot \lambda_{Ij,nw} \\
&= pw_I \cdot \exp\left(\alpha_Z + \beta_q' \cdot q_j + \beta_\$' \cdot \mu_{Ij,w} + \frac{1}{2} \cdot \beta_\$' \cdot \Omega_{Ij,w} \cdot \beta_\$\right) + \\
&+ \left(1 - pw_I\right) \cdot \left(\alpha_Z + \beta_q' \cdot q_j + \beta_\$' \cdot \mu_{Ij,nw} + \frac{1}{2} \cdot \beta_\$' \cdot \Omega_{Ij,nw} \cdot \beta_\$\right), \text{ with} \\
\beta_\$ &= \begin{bmatrix} \beta_p \\ \beta_m \end{bmatrix},
\end{aligned} \tag{10}
$$

where pw_I indicates the share of white residents in ZIP area I.[4] The exact derivation of income moments for a given ZIP and ethnic group is shown in Appendix A.

Based on preliminary specification tests and fit with the underlying data, we choose Cameron and Trivedi's [1] Negative Binomial II (Negbin II) specification as the specific count data framework to be used for this analysis. Using the result that the sum of a Negbin random variable follows the same distribution with first moment given by the sum of the individual expectations, the aggregate model to be estimated emerges as

$$
\begin{aligned}
p\left(D_{Ij} = Y_{Ij}\right) &= \frac{\Gamma\left(Y_{Ij} + N_I \cdot \nu\right)}{\Gamma\left(Y_{Ij} + 1\right) \cdot \Gamma\left(N_I \cdot \nu\right)} \cdot \left(\frac{\nu}{\lambda_{Ij} + \nu}\right)^{N_I \cdot \nu} \cdot \left(\frac{\lambda_{Ij}}{\lambda_{Ij} + \nu}\right)^{Y_{Ij}}, \text{ with} \\
D_{Ij} &= \sum_{i=1}^{N_I} d_{ij},
\end{aligned} \tag{11}
$$

where N_I indicates the total population of ZIP area I, λ_{Ij} is the representative individual's demand defined in (10), D_{Ij} is the (unobserved) total micro-demand from ZIP I to destination j, Γ denotes the mathematical gamma function, and ν is the index or precision parameter . The integer value Y_{Ij} indicates a realization of total visits per season from ZIP I to site j. The model in (11) is then estimated via QGPML. Appendix B shows the details of this process.

Application

The system demand model developed in the previous section is implemented using information on day trips to the Alpine Lakes Wilderness (ALW), which covers 393,000 acres in Washington's Mt. Baker-Snoqualmie National Forest. The area's proximity to some of the State's largest population centers makes it a popular hiking destination, with over 100,000 visitors per year. The numerous trails in the Wilderness are accessible through 51 trailheads, located around the Wilderness boundaries. Forty-nine of these "hiking zones" (HZs) are included in this analysis.

Permit data were provided by the National Forest Service.[5] The original set includes valid information on 14087 hiking groups, collected for the entire year of 1995. After filtering out multiple-day hikers and observations from ZIP codes that could not be matched with those included in the 1990 Census, 8750 valid observations on day-use group-visits from 134 different

ZIP code origins were retained for this analysis. The total number of residents in these zones is 3.01 million, roughly 55% of the State's population. Aggregating visits to a given HZ over ZIPs and introducing zero-visit values for the relevant ZIP / trail combination not represented by the original data yields a rectangular set of (134 x 49) = 6566 observations. These data are then combined with trail and access road information from an ALW guidebook (Spring et al. [22]), highway distances from a road atlas, and ZIP-specific information from the 1990 U.S. Census, as discussed in the previous section.

Estimation Results

We estimate both a simple aggregate Negbin (SANB) and a corrected aggregate Negbin (CANB) model. The simple model uses the naïve specification for λ_{Ij} (7) in the estimation of (11) while the corrected model applies (10) instead. Both models are estimated using QGPML.[6] Robust variance-covariance matrices for parameters are constructed applying the method suggested by White [25], and Gourieroux et al. [6]. This implies that SANB yields consistent estimates for coefficients and standard errors if (7) is the correct aggregate specification. The same is true for CANB with respect to (10). Naturally, since (7) is incorrect by assumption, the simple model will be flawed by aggregation bias, and bias resulting from using inconsistent estimates for λ_{Ij} in the Gourieroux et al. procedures. It is retained in this analysis merely for comparison purposes.

Table 1 presents parameter estimates for the two models. The components of q are trailhead elevation (measured in 1000 ft-units), a dummy taking the value of "1" if a lake can be reached within 5 miles of hiking from a given trailhead, and a dummy equal to "1" if the access road as defined above is longer than 5 miles. Travel cost are measured in dollars, and income in \$1000 units. The "PZ-" dummies correspond to α_Z described above. Both models fit the underlying data fairly well, with about two thirds of coefficient estimates significant at the 1 percent or 5 percent level. In general, the differences in parameter estimates and standard errors between the corrected and simple model are rather subtle. However, the corrected model produces a significantly lower and clearly more efficient parameter estimate for income, the variable subjected to the largest adjustments in the corrected version. To a smaller extent, this also holds for travel cost, which is a function of income as discussed above.

For both models, easy access to an alpine lake increases visitation rates by over 200 percent,[7] ceteris paribus. Also, visitors seem to have a slight preference for high-elevation trails based on the positive sign and significance of the corresponding coefficients. This result may be rooted in better views associated with higher elevations, and an increased chance for encountering alpine meadows, a generally positively valued trail feature (Englin and Shonkwiler [5]). In contrast, lengthy access roads are not as strong a deterrent as expected, based on the lacking significance of this coefficient. Perhaps this reflects the increased popularity of Sport-Utility Vehicles ("SUVs") and associated reduced problems in maneuvering these mostly unpaved secondary roads. The estimates for the different population zone dummies generally indicate that smaller, more rural communities have stronger preferences for day hiking in the ALW than larger population centers.

Beyond an inspection of coefficient estimates, the comparison of these models is somewhat problematic. Likelihood ratio tests are inapplicable since different regressors are used for the two specifications. Instead, we use goodness-of-fit statistics based on Pearson residuals, and deviance. These statistics are discussed in detail in Cameron and Windmeijer [2]. Based on these two measures, the corrected model performs slightly better than its simple counterpart.[8]

Estimates for elasticities and welfare measures for the two Negbin models are summarized in table 2. Elasticity measures are based on sample averages for travel cost and per-capita income, following e.g. Lutz et al. [17]. For CANB, they also reflect the average over the two race groups, weighted by population share. The own-price elasticity of demand for a "prototypical site" with price $\bar{p}$ is slightly greater than 1 for both models. The difference in elasticity estimates is more pronounced for income elasticity, which SANB estimates at 1.14, thirty percent higher than the values generated by CANB. The CANB figures of 0.85 to 0.88 seem to be more reasonable estimates considering the general finding in the literature of inelastic demand for recreation activities with respect to income (Loomis and Walsh [16]). For both elasticity groups, the corrected Nb model generates a tighter 95 percent confidence interval, as measured by the spread between lower and upper bound, divided by the mean estimate. This difference is especially pronounced for income elasticity, where the weighted spread for SANB is more than twice as large as the CANB counterpart for all residents (0.72 vs. 0.34).

Welfare estimates in table 2 are measured by compensating variation (CV), based on (5a).[9] Estimated per-trip welfare measures of $31-33 are consistent with recent findings in other studies on hiking behavior in Washington's Cascade mountains (e.g. Englin and Shonkwiler [5]). The corrected aggregate Negbin model produces a value of $1.83 million in total welfare generated by the ALW to the target population in 1995. The simple aggregate model over-estimates total welfare by approximately $90,000, or close to 5 percent. We use the simulation method suggested by Krinsky and Robb [13] to derive 95 percent confidence intervals for these values.[10] The results of this procedure mirror the findings for elasticities discussed above: The more efficient corrected model produces much tighter intervals. As measured by the spread-over-mean statistic, CANB outperforms SANB by 20 percent.

Sensitivity Analysis for Corrected Model

An inspection of (7) and (10) shows that the difference in welfare estimates between the two models largely depends on income effect, population shares, and income variance. To investigate the sensitivity of our results to model specifications, we perform a series of simulations with the same trip data, but altered population characteristics.[11] The results of these tests are captured in table 3. In the first simulation sequence (models 1a – 1e), we gradually reduce the proportion of white residents from an original ZIP average of 90 percent by up to 50 percent for all areas. This has a somewhat ambiguous effect on the bias of the simple model (measured as ratio of CVs / CVc, where *s* and *c* subscripts denote the simple and corrected model, respectively), but leads to a monotonic increase in efficiency, i.e. a progressively tighter confidence interval for CV generated by the corrected model, compared to the simple specification (last column of table 3). The second series (models 2a – 2e) represents a gradual increase in income variance from 10 percent to 300 percent. This leads to an opposite effect of monotonically increasing bias (from 5 percent to 10 percent), but ambiguous efficiency gains. The last test series, models 3a – 3d, combines a 20 percent decrease in the share of white residents with increases in income variance. Now both bias and efficiency loss of the simple model increase monotonically. Doubling income variance, for example, (model 3b) leads to an increase in bias of CV point estimates from 5 percent to 8 percent, and an increase in the relative width of the mean-calibrated confidence interval for CVs vs. CVc from 21 percent to a full 40 percent.

Such population proportions and income variances are not unrealistic, especially for larger urban population zones. This is illustrated in table 4, where the ZIP population and income characteristics of models 3a – 3d are compared to such information for five Metropolitan Statistical Areas (MSAs). As shown in column 3, the 20 percent reduction in whites across the 134 ZIPS used for this study leads to an average proportion of this group of around 72 percent, which compares well to analogous proportions for the MSAs. The remainder of the table shows averages and percentiles, over all ZIPs, of mean / variance ratios for income for both population groups. For example, a doubling of income variance for our data (model 3b) yields an average mean / variance ratio of 5 percent for white residents, and 11 percent for non-whites. Median values (at the 50^{th} percentile) are similar for this model. In comparison, such ratios correspond roughly to the lower quartile for ZIPs in MSAs. The upshot of this analysis is that a strong representation of recreation participants from zonal origins with racial diversity and large income variances in a given research project for could well lead to substantial aggregation bias and efficiency losses for welfare estimates in uncorrected zonal models.

Conclusion

This study shows that problems stemming from the use of aggregate data in zonal recreation models can affect both the theoretical relevance and analytical accuracy of estimation results. At a theoretical level, the parameter and welfare estimates generated by a misspecified aggregate model carry little informative value regarding the underlying recreation preferences of heterogeneous individuals. From an econometric perspective, the main drawback of an incomplete macro-specification is the loss of applicability of robust estimation techniques. This, in turn, exposes such models to additional sources of misspecification error.

In this paper we propose an extension to existing zonal count data models of recreation that captures some of the underlying heterogeneity of individual visitors by incorporating distributional information on per-capita income into the aggregate demand function. This adjustment eliminates the unrealistic constraint of constant income across zonal residents, and thus reduces the risk of aggregation bias in estimated macro-parameters. The suggested technique is easy to implement, and does not require the collection of additional data.

While relative bias and efficiency losses of welfare estimates generated by the uncorrected model are noticeable for the data used in this study, a follow-up analysis shows that the benefits of using the corrected specification increase with participation of residents from origin zones characterized by racial heterogeneity and high income variances.

Naturally, the enhanced count data framework suggested in this study will not completely eliminate aggregation problems in any given application. We believe, however, that our proposed adjustment strengthens the linkage between individual preferences and aggregate estimation for a wide spectrum of underlying micro-functions, and thus constitutes a general improvement over existing recreation models that rely on zonal data. Full recoverability of parameters for a micro-specification with fewer restrictions will almost certainly require additional information at the individual level.

Appendix A: Derivation of Income Variance from Census Information

In the 1990 U.S. Census, households are grouped into discrete income categories for each ZIP code. This information is also available in a nested version over race groups, although the number of income categories for this exposition is smaller than for the general grouping (9 vs. 25). Naturally, the lower bound for this discrete income distribution is given by zero. An upper bound, on the other hand, is not reported. The population mean and total for household income, $\overline{m}_{hh}$ and m_{tot} , are reported by the Census for any given race group. The remaining task is to estimate income variance.

As a first step, we estimate mean income in the highest category as

$$\overline{m}_K = \left(\frac{m_{tot} - \sum_{k=1}^{K-1} n_k \cdot \left(\frac{m_{\max,k} - m_{\min,k}}{2} \right)}{n_K} \right), \tag{A1}$$

where k is a category index, K indicates the highest income category, m_{tot} is total ZIP income, $m_{min,k}$ and $m_{max,k}$ are the lower and upper bounds of income category k, and n_k and n_K are the number of households in category k and the highest category, respectively. Race and ZIP-subscripts are omitted for convenience. The variance of household income can then be approximated by

$$\sigma^2_{hh} = \frac{\left(\sum_{k=1}^{K-1} n_k \cdot \left(\left(\frac{m_{\max,k} - m_{\min,k}}{2} \right) - \overline{m}_{hh} \right)^2 \right) + n_K \cdot \left(\overline{m}_K - \overline{m}_{hh} \right)^2}{N-1}, \tag{A2}$$

i.e. the average, over all households, of the sum of the squared deviations of bin midpoints (bin mean for the highest category) from the overall mean, with each such deviation weighted by the number of households in a given category.[12] This leads to the expression of the variance for per-capita income:

$$\sigma^2_{\mathrm{I,r}} = \left(\frac{1}{\overline{h}_{I,r}} \right)^2 \cdot \sigma^2_{\mathrm{hh,\,I,\,r}}, \quad \mathrm{r} = (\mathrm{w}, \mathrm{nw}), \tag{A3}$$

where $\overline{h}_{I,r}$ is the average household size for race group r in ZIP area I.

Appendix B: Model Estimation Through Generalized Maximum Likelihood

A correct specification of the aggregate demand function notwithstanding, parameter estimates can still be biased if the probability density function for d_{ij} is misspecified (Gourieroux et al. [6] and [7]). Specifically, if the distributional assumptions reflected in (11) do not hold for the Negbin model, ML estimates will be inconsistent. PML and QGPML methods guard against such specification error. For the Negbin model, either of the two estimation techniques can be used to derive consistent coefficients and robust standard errors, but QGPML estimators are generally more efficient. Both methods differ from ML by arbitrarily specifying a value for v in (11) instead of treating it as an additional model parameter. This reduces the relevant part of the LLF to

$$ll_{(Nb)} = \sum_{I=1}^{I_T}\sum_{j=1}^{J}\left[a\cdot\left(\ln(a)-\ln\left(\Lambda_{Ij}+a\right)\right)+Y_{Ij}\cdot\left(\ln\left(\Lambda_{Ij}\right)-\ln\left(\Lambda_{Ij}+a\right)\right)\right], \quad \text{where} \qquad \text{(B1)}$$

$$\Lambda_{Ij} = N_I \cdot \lambda_{Ij}$$

Following the notation in Gourieroux et al. [7], a represents the substitute for $(N_I \bullet v)$. The QGPML process requires two additional estimation steps. In an interim procedure, a consistent estimate for $V = N_I \bullet v$ is derived by using the Negbin II expression for the variance of D_{Ij} and the parameter estimates from maximization of (B1) in the following regression:

$$\left(Y_{Ij}-\hat{\Lambda}_{Ij}\right)^2-\hat{\Lambda}_{Ij} = \frac{1}{V}\cdot\hat{\Lambda}_{Ij}^{\,2} \qquad \text{(B2)}$$

The OLS estimate for V in (B2), $\hat{V}$, is then inserted into (B1) in lieu of a. A second round of optimization using this adjusted LLF generates the final results.

Notes

[1] The theoretical underpinnings to these concepts are discussed in Gourieroux et al. [6]. In a related study [7] the same authors show how PML and QGPML estimation can be used in the context of count data models. The general idea behind this approach is appealing: If the dependent variable, say trips to a recreation site, is characterized by a probability distribution that is a member of the linear exponential family, such as Poisson or Negbin, ML estimation yields consistent parameter estimates, even if the empirical model does not reflect the true distribution of site demand.

[2] Within the framework of a count data model actual demands by individual i for site j are assumed to be unobserved. Site demand d_{ij} is treated as a random variable with an expected value commonly labeled as λ_{ij}.

[3] Source: U.S. DoT Web Site: http://www.bts.gov/btsprod/nts/chp2/tbl2x18.html. The $0.25/mile figure constitutes a compromise between $0.41 for total, and $0.1 for variable costs (1995).

[4] Naturally, this specification implies that the proportions of race groups among visitors are equal to racial shares in the general ZIP population at all destinations. If this assumption seems implausible in a given application, equation (10) can easily be modified to allow for population shares specific to a given origin – destination pair, if empirical estimates or "educated guesses" for these shares are available. This generalization would be captured in (10) by assigning "*Ij*" subscripts to *pw*, and *pnw*. All other population parameters would be still based on ZIP code characteristics, and remain unchanged.

[5] We are indebted to Gary Paull, Wilderness trail coordinator, and his colleagues at the Mt. Baker-Snoqualmie NF headquarters in Mt. Lake Terrace, WA, for provision of data and other helpful information.

[6] We use Matlab's Trust Region Method with user-supplied gradients and Hessian matrices as optimization algorithm for this analysis. The full Matlab program is available from the author upon request. We thank Daniel Hellerstein for providing count data GAUSS code, which proved very useful in developing this program.

[7] We use Halvorsen and Palmquist's [8] formula of $(exp(\beta) - 1)$as an approximation for this marginal effect.

[8] The R-squared values for CANB and SANB are 0.306 vs. 0.271 based on Pearson residuals, and 0.348 and 0.335 based on deviance.

[9] We also computed estimates for consumer surplus (CS) and equivalent variation (EV). The three welfare measures differ only marginally from each other within a specific aggregate model. This is an expected result, given the small income effects mentioned earlier. The relatively insignificant magnitude of these divergences notwithstanding, the three measures are still ranked as required by utility-theory for an increase in prices, with $|CV| > |CS| > |EV|$ for both models.

[10] Specifically, we draw 1000 sets of coefficients, based on parameter estimates and their variance-covariance matrix. For each set, we then derive a corresponding set of λ_{Ij} for all Ij combinations. This leads to the computation of CV for each ZIP code area, using (5a) weighted by ZIP population. These values are added over all ZIPs to yield 1000 estimates of total CV. The reported confidence intervals are based on the empirical distribution of these estimates. Elasticities, on the other hand, are linear in parameters for our model, so the derivation of their standard deviation and corresponding confidence intervals is straightforward.

[11] These simulations are based on the assumption that the alterations in population characteristics would not change observed trip behavior. This assumption seems tenable, given that mean income remains unchanged for all ZIPs, and income effects are generally small.

[12] This variance estimator is slightly biased upwards, but generally performed well in Monte Carlo simulations. A possible extension for the computation of $\sigma^2_{I,r}$ would be to apply more advanced nonparametric techniques, such as the histospline - method proposed by Minnotte [18].

References

Cameron, A. Colin, and Pravin K. Trivedi, Econometric models based on count data: Comparison and applications of some estimators and tests, *Journal of Applied Econometrics* **1**, 29-53 (1986).

Cameron, A. Colin, and Frank A. G. Windmeijer, R-Squared measures for count data regression models with applications to health-care utilization, *Journal of Business & Economic Statistics* **14**, 209-220 (1996).

Cicchetti, Charles J., Anthony C. Fisher, and V. Kerry Smith, An econometric evaluation of a generalized consumer surplus measure: The Mineral King controversy, *Econometrica* **44**, 1259-1276 (1976).

Englin, Jeffrey, Peter Boxall, and David Watson, Modeling recreation demand in a Poisson system of equations: an analysis of the impact of international exchange rates, *American Journal of Agricultural Economics* **80**, 255-263 (1998).

Englin, Jeffrey, and J. S. Shonkwiler, Estimating social welfare using count data models: An application to the long-run recreation demand under conditions of endogenous stratification and truncation., *The Review of Economics and Statistics* **77**, 104-12 (1995).

Gourieroux, C., A. Monfort, and A. Trognon, Pseudo maximum likelihood methods: theory, *Econometrica* **52**, 681-700 (1984a).

Gourieroux, C., A. Monfort, and A. Trognon, Pseudo maximum likelihood methods: application to Poisson models, *Econometrica* **52**, 701-720 (1984b).

Halvorsen, Robert, and Raymond Palmquist, The interpretation of dummy variables in semi-logarithmic equations, *American Economic Review* **70**, 474-475 (1980).

Hellerstein, Daniel, Using count data models in travel cost analysis with aggregate data, *American Journal of Ariculturel Economics* **73**, 860-867 (1991).

Hellerstein, Daniel, The treatment of nonparticipants in travel cost analysis and other demand models, *Water Resources Research* **28**, 1999-2004 (1992).

Hellerstein, Daniel, Welfare estimation using aggregate and individual-observation models: a comparison using Monte Carlo techniques, *American Journal of Agricultural Economics* **77**, 620-630 (1995).

Hilger, James, and Jeffrey Englin, Utility theoretic linear exponential demand systems: an analysis of forest fire impacts on recreational values and uses, *Working paper*, Department of Applied Economics and Statistics, University of Nevada, Reno (2000).

Krinsky, Itzhak, and A. Leslie Robb, On approximating the statistical properties of elasticities, *Review of Economics and Statistics* **68**, 715-719 (1986).

LaFrance, Jeffrey T., Incomplete demand systems and semilogarithmic demand models, *Australian Journal of Agricultural Economics* **34**, 118-131 (1990).

LaFrance, Jeffrey T., and W. Michael Hanemann, The dual structure of incomplete demand systems, *American Journal of Agricultural Economics* **71**, 262-274 (1989).

Loomis, John B., and Richard G. Walsh. *Recreation economics decisions - comparing benefits and costs*. 2nd edition ed. State College, PA: Venture Publishing, Inc. (1997).

Lutz, Janet, Jeffrey Englin, and J. Scott Shonkwiler, On the aggregate value of recreational activities - A nested price index approach using Poisson demand systems, *Environmental and Resource Economics* **15**, 217-226 (2000).

Minnotte, Michael C., Achieving higher-order convergence rates for density estimation with binned data, *Journal of the American Statistical Association* **93**, 663-672 (1998).

Shaw, Daigee, On-site samples' regression - problems of non-negative integers, truncation, and endogenous stratification, *Journal of Econometrics* **37**, 211-223 (1988).

Shonkwiler, J. S. Recreation demand systems for multiple site count data travel cost models, *in* "Valuing recreation and the environment: revealed preference methods in theory and practice", (J. A. Herriges and C. L. Kling, Ed.). Cheltenham, U.K., and Northampton MA: Edward Elgar (1999).

Shonkwiler, J. S., and Jeffrey Englin, Welfare losses due to livestock grazing on public lands: some evidence from the Hoover Wilderness, *Paper presented at the W-133 Meetings in Tucson, Arizona* (1999).

Spring, Vicky, Ira Spring, and Harvey Manning. *100 Hikes in Washington's Alpine Lakes Wilderness*. 2nd ed. Seattle: The Mountaineers (1993).

Stoker, Thomas M., Completeness, distribution restrictions, and the form of aggregate functions, *Econometrica* **52**, 887-908 (1984).

Stoker, Thomas M., Empirical approaches to the problem of aggregation over individuals, *Journal of Economic Literature* **31**, 1827-1874 (1993).

White, Halbert, Maximum likelihood estimation of misspecified models, *Econometrica* **50**, 1-25 (1982).

Zellner, Arnold, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *American Statistical Association Journal* 351-368 (1962).

Table 1: Estimation Results

a (b) = sign. at 1% (5%)	SANB		CANB	
Variable	coeff.	s.e	coeff.	s.e
Intercept	-9.630	0.278 [a]	-9.427	0.194 [a]
Trailhead elevation	0.126	0.039 [a]	0.144	0.051 [a]
Lake w/in 5 miles	1.152	0.075 [a]	1.119	0.078 [a]
Acces road > 5 miles	-0.060	0.072	-0.001	0.083
Travel cost	-0.031	0.004 [a]	-0.032	0.003 [a]
Income	0.060	0.011 [a]	0.045	0.004 [a]
PZ 2	-0.217	0.172	-0.217	0.168
PZ 3	0.815	0.188 [a]	0.827	0.184 [a]
PZ 4	0.149	0.206	0.159	0.204
PZ 5	1.146	0.266 [a]	1.137	0.251 [a]
PZ 6	1.566	0.289 [a]	1.528	0.280 [a]
PZ 7	0.614	0.203 [a]	0.593	0.175 [a]
PZ 8	0.326	0.184	0.368	0.185 [b]
PZ 9	-0.590	0.190 [a]	-0.619	0.204 [a]
PZ 10	2.360	0.321 [a]	2.226	0.309 [a]
PZ 11	1.250	0.388 [a]	1.098	0.380 [a]
PZ 12	0.806	0.273 [a]	0.685	0.273 [b]
PZ 13	2.203	0.410 [a]	2.094	0.390 [a]
PZ 14	1.970	0.415 [a]	1.828	0.397 [a]
PZ 15	-0.212	0.269	-0.141	0.266
PZ 16	-0.791	0.266 [a]	-0.845	0.261 [a]
PZ 17	0.344	0.388	0.448	0.408
PZ 18	0.304	0.267	0.196	0.259
PZ 19	2.123	0.289 [a]	2.081	0.297 [a]
PZ 20	-0.297	0.278	-0.225	0.372
PZ 21	0.719	0.369 [b]	0.669	0.347
PZ 22	-1.355	0.370 [a]	-1.334	0.388 [a]
PZ 23	0.886	0.366 [b]	0.905	0.364 [b]
PZ 24	1.187	0.549 [b]	1.228	0.609 [b]
PZ 25	-0.372	0.264	-0.378	0.276
PZ 26	0.909	0.356 [b]	0.944	0.345 [a]
PZ 27	-0.978	0.514	-0.926	0.524
PZ 28	2.144	0.467 [a]	1.996	0.425 [a]
PZ 29	0.125	0.511	0.086	0.438
PZ 30	-1.446	0.495 [a]	-1.443	0.501 [a]
PZ 31	0.047	0.389	-0.082	0.381
PZ 32	1.241	0.161 [a]	1.198	0.156 [a]
PZ 33	0.991	0.160 [a]	0.853	0.154 [a]
PZ 34	0.340	0.159 [b]	0.329	0.155 [b]
PZ 35	0.752	0.187 [a]	0.751	0.180 [a]

Table 2: Elasticity and Welfare Estimates

Statistic	SANB	CANB
Elasticities		
Price		
L.B.	-0.91	-1.06
Estimate	-1.23	-1.29
U.B.	-1.54	-1.53
Spread over mean	*0.51*	*0.37*
Income		
L.B.	0.73	0.73
Estimate	1.14	0.88
U.B.	1.55	1.03
Spread over mean	*0.72*	*0.34*
CV		
Per trip (ZIP average)	32.77	31.07
Total, all ZIPs		
L.B.	1,587,808	1,599,931
Estimate	1,922,749	1,833,440
U.B.	2,510,312	2,324,822
Spread over mean	*0.48*	*0.40*

L.B. (U.B.) = lower bound (upper bound) for 95 percent confidence interval
For computation of confidence intervals see footnote 10.

Table 3: Model Simulations with Different Population Shares and Income Variances

Model	Simulations		CVs	CVc	Ratio CVs / CVc	Ratio SOMs /SOMc
	Redution in share of white population	Increase in income variance (both groups)				
original	0%	0%	1,922,774	1,833,373	1.049	1.213
1a	10%	0%	1,927,901	1,828,717	1.054	1.256
1b	20%	0%	1,920,959	1,828,384	1.051	1.276
1c	30%	0%	1,942,608	1,830,410	1.061	1.305
1d	40%	0%	1,954,637	1,834,008	1.066	1.307
1e	50%	0%	1,962,951	1,839,131	1.067	1.330
original	0%	0%	1,922,774	1,833,373	1.049	1.213
2a	0%	10%	1,922,774	1,828,744	1.051	1.235
2b	0%	30%	1,922,774	1,819,740	1.057	1.091
2c	0%	50%	1,922,774	1,810,651	1.062	1.191
2d	0%	200%	1,922,774	1,789,074	1.075	1.346
2e	0%	300%	1,922,774	1,743,421	1.103	1.472
original	0%	0%	1,922,774	1,833,373	1.049	1.213
3a	20%	50%	1,920,959	1,801,645	1.066	1.377
3b	20%	100%	1,920,959	1,776,244	1.081	1.400
3c	20%	150%	1,920,959	1,750,379	1.097	1.577
3d	20%	200%	1,920,959	1,725,935	1.113	1.650

CVs = CV for simple model
CVc = CV for corrected model
SoMs = Spread of confidence interval of CV over mean of CV for simple model
SoMc = Spread of confidence interval of CV over mean of CV for corrected model

Table 4: Distribution of Mean-Variance Ratios for Income over ZIPs

Model / Area	No. of zips	pw	Mean-Variance Ratio for Income* Average w	nw	Percentiles 10 w	10 nw	25 w	25 nw	50 w	50 nw	75 w	75 nw
original	134	0.90	0.10	0.21	0.05	0.07	0.08	0.12	0.10	0.15	0.13	0.22
3a	134	0.72	0.07	0.14	0.03	0.05	0.05	0.08	0.07	0.10	0.09	0.14
3b	134	0.72	0.05	0.11	0.02	0.04	0.04	0.06	0.05	0.08	0.06	0.11
3c	134	0.72	0.04	0.08	0.02	0.03	0.03	0.05	0.04	0.06	0.05	0.09
3d	134	0.72	0.03	0.07	0.02	0.02	0.03	0.04	0.03	0.05	0.04	0.07
L.A.	273	0.63	0.13	0.19	0.03	0.06	0.05	0.10	0.09	0.17	0.16	0.27
Bay Area	228	0.72	0.10	0.17	0.03	0.07	0.05	0.10	0.08	0.14	0.13	0.21
California	1495	0.77	0.14	0.26	0.04	0.07	0.07	0.13	0.11	0.23	0.19	0.35
Chicago	202	0.78	0.13	0.19	0.04	0.07	0.07	0.10	0.12	0.18	0.17	0.24
Miami	78	0.75	0.14	0.26	0.03	0.095	0.05	0.16	0.12	0.26	0.20	0.34

pw = percentage of white population (ZIP average)
w = white, nw = non-white
*Income measured in $000

Single Bounded versus Multiple Bounded Estimators: Some Evidence on Efficiency Gains

Daniel Hellerstein
ERS/USDA

and

Kevin Boyle
University of Maine, Orono

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

**The view expressed in this paper are those of the authors alone, and do not necessarily represent policies or views of their respective institutions.

Abstract: In this paper we consider the multiple bounded extension to single bounded discrete choice contingent valuation. Using a payment-card like format, multiple bounded techniques should improve estimator efficiency. This possibility is investigated using single bounded and multiple bounded estimators applied to several variants of a survey on the protection of Iowa grassland birds. In addition, to account for non-participation and censoring, two stage estimators are introduced, and tested against both observed and simulated data.

In the course of human events, is often necessary to evaluate programs that deliver quantities of environmental goods and services, to produce a money based metric of their value. As is well known, the public goods aspects of birds, bees, flowers, trees, etc. complicate the measurement of these benefits. Hence, measures based on indirect techniques are often required.

Stated preference methods are amongst the more important of these indirect techniques. Here, the public is engaged in a structured conversation, the purpose of which is to ferret out their true desires. In this work, we consider an efficiency consideration regarding the applications of the contingent valuation method. Specifically, the use of multiple valuation questions as opposed to a single query.

Let's start with a methodological conundrum: survey respondents should not be forced into uncomfortable levels of cogitation when faced with a valuation question, yet analysis is best served by exactitude when obtaining the respondent's willingness to pay for the desired environmental good. That is: tempering the investigator's need to intensively probe for the respondent's real attitudes is the necessity of engaging a not-necessarily-highly-motivated public to bother responding at all.

It is thought by many that an opened ended query runs afoul of this quandary, for it demands too much of the average citizen [Desvousges et al, Bishop and Heberlein]. Instead, a discrete-choice approach is oft advocated [Arrow et al, Haneman and Kanninen]: when presented with a hypothetical price, the respondent answers whether she would choose to purchase (say, via tax or contribution) this good (or service). Such a "single bounded" valuation exercise shares much with normal purchasing decisions, and ought to be intuitively simpler for respondents to consider.

Yet this goal of making it easy on the respondent makes it harder on the analyst; since rather then a explicitly stated value one must work with a discrete choice at a single value, and each of the finite set of choices (for example, a "yes" or a "no") coincide with a wide range of underlying values held by a respondent. As it lacks the leverage found in continuous models, coefficient accuracy will be degraded (that is, standard errors will increase), with concomitant effects on the accuracy of WTP measures [Haneman].

In order to improve upon efficient while retaining the ease of discrete choice, one may use multiple bounded models. For example, the double bounded approach [Haneman, Loomis and Kanninen] asks respondents a follow up question that attempts to bracket the respondent's value.[1] Extending this logic, what if a set of follow up questions were asked, with the goal of obtaining a narrow gap between the respondent's upper and lower bounds? A multiple bounded survey format [Welsh and Poe] is one way of implementing such a series of questions. The respondent is presented with a sequence of values (say, ranging from $1 to $100), and is asked to "circle" the highest value he would pay. Implicitly, this is the upper bound (V+dv), while the next lower value is the lower bound (V).

Although appealing in a pure sense, the additional complications of a multiple bounded approach may outweigh any efficiency gain. For example, double bounded models may be prone to yea saying and other forms of respondent fatigue, with later responses based on survey

[1] For example, if the respondent answers "YES" to a value of V, then the follow up question may ask for an opinion on V+dv. If the answer is NO to this follow up, then the analyst knows that the respondent's value lies between V and V+dv.

structure rather then intrinsic values. In the multiple bounded case, the order of presentation may matter, as may the range of values used [Alberini et al].

In this paper, we compare the performance of single bounded and multiple bounded estimators. Using data on the value of grassland birds, the results of a multiple bounded survey are compared to a single bounded survey. Of particular interest is a comparison of the capability of identifying scope effects. In addition, we consider alternative means of handling non-participation, using censored estimators and a two-stage estimator, using both the observed data and simulations.

The Data

In order to minimize anchoring bias, and other problems of intra personal comparisons, we use the results of several separate contingent valuation surveys (each using a different sample of respondents) that looked at the value of increasing the population of grassland birds in Iowa [Boyle et al]. We focus on four variants, comprising two different levels of protection, and two different response modes. These are:

SB_S: Single bounded question, "small change" in bird numbers
SB_L: Single bounded question, "large change" in bird numbers
MB_S: Multiple bounded questions, "small change" in bird numbers
MB_L: Multiple bounded questions, "large change" in bird numbers

Since scope effects [Diamond and Hausman; Carson] are of especial interest, the single bounded and multiple bounded variants were combined, with a %CHANGE variable added to account for the size of the change in bird numbers.

The single bounded variants asked respondents to provide a Yes, No or Don't Know response to a one time increase in Iowa income tax to provide for modifications to CRP land that would increase the population of several species of grassland birds. A follow up question was then asked regarding how certain they were -- ranging from "definitely yes", to "definitely no".

In contrast, the multiple bounded variants presented respondents with a 12 row by 5 column answer sheet. Each row contained a value (ranging from $1 to $100), and each column contained a <u>certainty</u> level (from "Definitely Yes" to "Definitely No"). For each cost value, respondent's circled the appropriate <u>certainty</u> level.[2]

In all other aspects, the variants were essentially the same; they were drawn from the same population (Iowa residents), the descriptive material and debriefing questions were the same, and the scenarios were the same within a variant (the small change variants posit an 8% increase in several bird species, the large change variants posit a 100% increase).

In addition, each variant preceded the valuation question with a "would you vote for the proposal if it would increase your household's income tax". A NO answer identifies "non-participants", who were not interested in the proposed change in bird population. As shown below, how one treats these respondents can influence estimated results.

Figure A shows how responses vary across program size and response mode. As expected, the multiple bounded variants show an clear increase in the probability of saying NO (at any bid), and show a slight but consistent increase as the program size diminishes. The

[2] For example, a respondent might circle "Definitely Yes" for values of $1, $2, and $5; then "Probably Yes" for values of $7 and $10, and "Definitely No" for all other values.

results for the single bounded model are much less definitive, with the probability of NO only loosely correlated with program size, and showing a much weaker responsiveness to bid value.[3]

Although suggestive of the greater power of multiple bounded models, a parametric approach should yield more definitive answers. In particular, we now look at several single bounded and multiple bounded probit estimators.

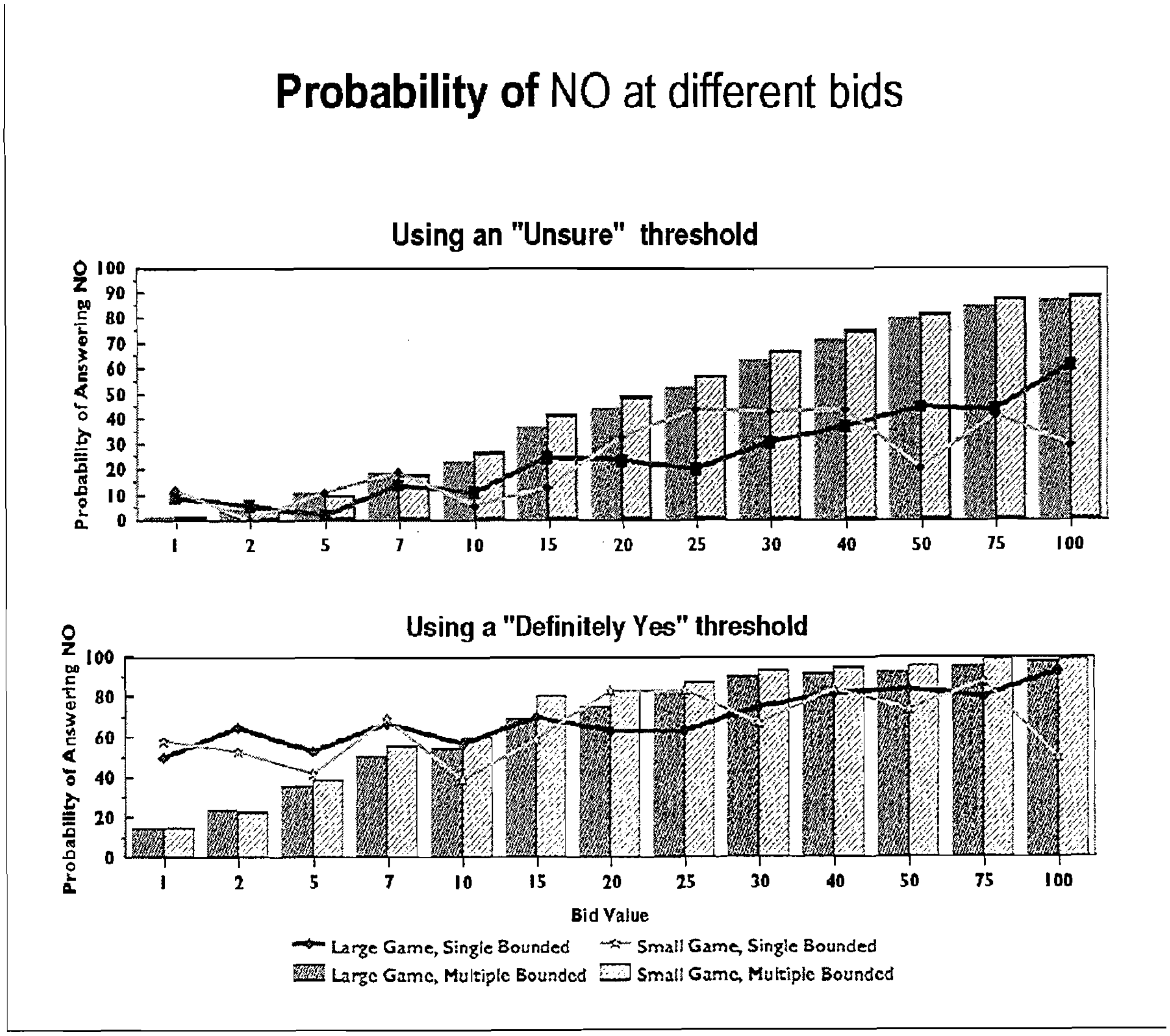


Figure A.

[3] The multiple bounded charts use cumulative results, hence are guaranteed to increase. In contrast the single bounded charts the probability at any bid value is based on respondents offered that bid. Hence, single bounded probabilities may decrease as bid value increases.

Single Bounded, Multiple Bounded, and Bivariate Probit

Both multiple bounded and single bounded estimators assume that an individual will answer YES when her willingness to pay for the change (say, 100% more birds) exceeds the offered bid. Per standard practice, we assume:

$$WTP = \mu + \varepsilon \ \ ; where\ \varepsilon\_Normal(0,\sigma)$$

and, when WTP exceeds (is less then) the offered bid, then the respondent will answer YES (NO).

The single bounded probit estimator has a likelihood of:

$$\_ = \prod_{y=1} \Phi(\mu) * \prod_{y=0} (1 - \Phi(\mu))$$

$$where$$

$$(1) \quad y = 1\,(0) \ : \ YES\ (NO)\ respondents$$

$$\Phi = Standard\ normal\ CDF$$

$$\mu = \frac{X\beta\ \text{-}\ bid}{\sigma}$$

Equation 1 is maximized with respect to β and σ. Note that, in contrast to the standard probit, σ is estimable hence β can be identified and absolute willingness to pay (WTP) measures can be computed.

The multiple bounded probit estimator [Welsh et al] has a likelihood of:

$$\_ = \prod \Phi(\mu_u) - \Phi(\mu_l)$$

$$where$$

$$\mu_u = \frac{bid_u - X\beta}{\sigma}$$

$$(2) \qquad \mu_l = \frac{bid_l - X\beta}{\sigma}$$

$$bid_l = r'espondents\ lower\ bound$$

$$bid_u = r'espondents\ upper\text{-}bound$$

As with equation 1, maximization is with respect to β and σ.

When the respondent always answers NO (at all bid levels), an appropriate lower bound must be assigned.[4] There are several possibilities:

a) assign a lower bound of 0

Solution a forces the respondent to have a strictly positive WTP.

b) assign a lower bound of $-\infty$

Solution b is analytically equivalent to a "spike" [Kristom] model, that assumes that respondent's have a probability spike at a net value of 0, with:

i) Prob $\varepsilon=-X\beta$ is equal to $\Phi(-X\beta/\sigma)$, and

ii) Prob $\varepsilon<-X\beta$ is equal to 0

c) truncate the distribution at $-X\beta$

Solution c is similar to solution a, but corrects the normal distribution to yield a proper distribution (with a CDF equal to 1 and +infinity).

In addition to "all NO" answers, a substantial fraction (about 20%) were "non-participants", they stated that they would not pay anything for a variant. There are several ways to incorporate these individuals:

a) Discard them

If WTP values are to be aggregated to the entire population, one must adjust for this non-participation rate.

b) Assume they have values between -infinity and 0.

In contrast, individuals who are participants, but say NO at the lowest offered bid, will be assigned bounds between 0 and this lowest bid. For the single bounded model, the appropriate bid value would be zero.

c) Estimate a two-stage model

The first stage examines the probability of participation, the second the WTP given participation.

The two stage model can be constructed to allow a correlation between stages. In particular, a bivariate normal distribution can be used to estimate a two-stage single bounded and a two-state multiple bounded model. As with the single bounded probit, we have:

i) Participation occurs if $\mu_1 + \varepsilon_1 > 0$.

ii) A YES response at bid B occurs if $\mu_2 + \varepsilon_2 > B$.

Note that, as with the Heckman linear model (Maddala), we allow each stage to have it's own μ and ε. This implies that each stage may be a function of a unique set of covariates.

The two stage single bounded probit estimator [Greene] has a likelihood of:

[4] When the respondent always answers YES, an upper bound must be assigned (say, a very large positive value). This is less of a concern, since the influence on the likelihood functions of one's upper bound is small; few individuals always answer yes, and these individuals tend to have large positive ε (that is, estimated μ is often much less then the highest bid).

$$\_ = \prod_{y1=0} 1-\Phi(\mu_1) * \prod_{y1=1,y2=0} \Phi_2(\mu_1,-\mu_2,-\rho) * \prod_{y1=1,y2=1} \Phi_2(\mu_1,\mu_2,\rho)$$

where

Φ_2= *bivariate normal cdf*

$$\mu_1 = X_1\beta_1$$

$$\mu_2 = \frac{X_2\beta_2 - bid}{\sigma}$$

X_1, β_1 = *participation stage covariates and coefficients*

X_2, β_2 = *bid response stage covariates and coefficients*

ρ = *correlation coefficient*

Note that, as is typical of probit estimators, the first stage coefficients are identified up to a scale constant (that is, the first stage σ is assumed to equal 1.0).

The two stage multiple bounded probit estimator has a likelihood of:

$$\_ = \prod_{y1=0} 1-\Phi(\mu_1) * \prod_{y1=1} \Phi_2(\mu_1,\mu_{2l},\rho) - \Phi_2(\mu_1,\mu_{2u},\rho)$$

where

Φ_2= *bivariate normal cdf*

$$\mu_1 = X_1\beta_1$$

$$\mu_{2s} = \frac{X_2\beta_2 - bid_s}{\sigma}.\ bid_l = lower\ bound;\ bid_u = upper\ bound$$

X_1, β_1 = *participation stage covariates and coefficients*

X_2, β_2 = *bid response stage covariates and coefficients*

ρ = *correlation coefficient*

Note that with the two stage multiple bounded estimator, respondents who answer NO at the lowest offered bid ($1 in this case) are assumed to have a lower bound of $0 and an upper bound of $1.[5]

[5] GAUSS software to compute these several estimators is freely available from the authors upon request.

Willingness to pay is computed using both "censored" and "hanemann" [Haneman and Kanninen] formulae.

Censored WTP:

$$\textit{Single bounded}: \Phi(X\beta/\sigma)*X\beta + \sigma\phi(X\beta/\sigma)$$

$$\textit{Double bounded}: \Phi_1 X_2 \beta_2 + \rho\sigma\phi_1$$

$$\textit{where}:$$

$$\Phi_1 = \Phi(X_1\beta_1)$$

$$\phi_1 = \phi(X_1\beta_1)$$

and

Haneman WTP:

$$\textit{Single bounded}: \int_0^{\textit{infinity}} \left[1 - \Phi\left(\tfrac{x - X\beta}{\sigma}\right)\right]dx$$

$$\textit{Double bounded}: \int_{-\mu_1}^{\textit{infinity}} \phi(x-\mu_1)\left[(\mathrm{M}_x + (\mathrm{T}(-\mathrm{M}_x/\Sigma)*\Sigma)\right]dx$$

$$\textit{where}:$$

$$\Sigma = \sigma\sqrt{(1-\rho^2)}\ ;\ \textit{conditional sd of } X_2$$

$$\mathrm{M}_x = X_2\beta_2 + \rho\sigma x_i\ ;\ \textit{conditional mean of } X_2|x$$

$$\mathrm{T}(z) = \phi(z)/(1-\Phi(z));\ \textit{the mills ratio}$$

$$\mu_1 = X_1\beta_1$$

$$\textit{Note that } (\mathrm{T}(-\mathrm{M}_x/\Sigma)*\Sigma)\ \textit{is the expected value of } \varepsilon_2|\varepsilon_2 > -X_2\beta_2$$

Note that the two stage Haneman formula is a transformation of the area under a bivariate normal CDF.

Model Results

Table I and II shows the results for several estimators using pooled (across program size) Iowa data. Table I lists a parsimonious model, whereas table II lists models that include a few socioeconomic covariates.[6]

[6] The respondent's household income, and a dummy indicating whether anyone in the household hunts or views birds. A number of other specifications, using attitudinal, education, and other variables were also estimated. These yielded qualitatively similar results.

Table 1a. Iowa Results, simple model

Model	Probit	SBP, all	SBP, participants	MBP, all	MBP, participants
Constant	0.53 (7.0)	-772.1 (-0.45)	-14.5 (-1.2)	0.842 (1.1)	8.87 (9.7)
BirdPct	0.00187 (1.7)	-0.719 (-0.35)	-0.126 (-0.99)	0.046 (3.5)	0.047 (4.1)
σ	n.a.	1152.2 (0.46)	99.6 (4.5)	23.7 (40.3)	20.76 (41.3)
Wald Test (Prob(β=0))	163.3 (1.7e-38)	0.338 (0.85)	41.1 (1.8e-09)	1730 (0)	1716 (0)
# obs	753	766	548	1531	1131
WTP, 8% WTP, 100%	n.a.	172.00 156.00	32.4 27.6	10.1 12.4	13.7 16.9

Notes:

- SBP is the single bounded probit. MBP is the multiple bounded probit.
- "All" includes all respondents, with "non-participants" either assigned a NO bid at $0, or assigned bounds between -infinity and 0.
- "Participants" variants only include those who would pay something.
- WTP is computed using the "censored" WTP.
- WTP, 8% is the average willingness to pay (across all respondents) for an 8% increase.
- WTP, 100% is the average willingness to pay (across all respondents) for an 100% increase.

Table 1b. Iowa Results, simple model. Two stage models

		Two Stage SBP	Two Stage MBP
First (participation) stage	Constant	0.525 (6.3)	0.193 (3.3)
	BirdPct	4.13e-5 (0.036)	0.00195 (2.58)
Second (bid response) stage	Constant	39.7 (1.55)	3.64 (3.39)
	BirdPct	-0.0055 (-0.07)	0.040 (3.1)
	σ	98.7 (2.7)	20.5 (78.8)
	ρ	-0.82 (-1.428)	0.999 (n.a.)
Wald Test		197.1 (1.2e-40)	16169 (0)
# obs		568	1531
WTP, 8% WTP, 100%		35.3 35.3	10.2 12.4

Notes:
- WTP is the two-stage Haneman WTP.
- ρ is the correlation coefficient, and is constrained to lie between -1 and 1. When ρ is at either constraint, the standard error is not estimated.

Table 2a. Iowa Results

Model	Probit	SBP	SBP, participants	MBP	MBP, participants
Constant	0.040 (0.30)	-663.0 (-0.82)	-47.5 (-2.04)	-11.9 (-6.9)	-1.49 (-1.53)
BirdPct	0.00225 (1.97)	-0.382 (-0.46)	-0.119 (-0.92)	0.051 (3.6)	0.051 (3.97)
Income	5.89e-06 (2.9)	0.0031 (0.82)	3.67e-4 (1.7)	1.5e-4 (6.8)	1.3e-4 (6.8)
HuntView	0.414 (3.8)	225.6 (0.83)	27.1 (1.89)	10.5 (7.3)	6.74 (5.4)
σ	n.a.	601.3 (0.84)	99.1 (4.45)	23.1 (38.0)	20.5 (39.9)
Wald Test (Prob(β=0))	184.6 (8.7e-40)	1.16 (0.88)	39.4 (5.6e-8)	1556.4 (0)	1807.8 (0)
# obs	693	707	515	1377	1041
WTP, 8% WTP, 100%	n.a.	101.00 91.8	34.0 29.3	10.7 13.4	14.2 17.6

Notes:
- WTP computed using the "censored" WTP.
- Income is yearly household income.
- HuntView is a 0/1 dummy, that is equal to 1 if someone in the house either hunted or viewed wildlife.

Table 2b. Iowa Results. Two stage models

		Two Stage SBP	Two Stage MBP
First (participation) stage	Constant	-0.064 (-0.45)	-0.358 (-3.5)
	Birdpct	5.9e-4 (0.48)	0.0023 (2.7)
	Income	2.4e-6 (1.1)	6.4e-6 (5.8)
	HuntView	0.463 (3.9)	0.44 (5.2)
Second (bid response) stage	Constant	0.975 (0.049)	-7.74 (-4.2)
	Birdpct	-0.053 (-0.58)	0.047 (3.3)
	Income	1.3e-4 (1.02)	1.3e-4 (8.6)
	Huntview	33.2 (3.3)	9.11 (6.1)
	σ	89.5 (5.7)	20.5 (69)
	ρ	-0.70 (-2.2)	0.99 (n.a.)
Wald Test		118.6 (2.4e-21)	14364 (0)
# obs		532	1368
WTP, 8% WTP, 100%		36.1 16.3	10.9 13.6

Notes:

- WTP is the two-stage Haneman WTP.

The most striking feature is the poor performance of the single bounded probit model that uses all observations (SBP, all). This is probably due to the abundance of "NO at $0" observations (the imputed values for respondents who answered no to the 'participation' question). The SBP model of participants is better, with significant Wald statistic. However, each model predicts a disturbingly negative (albeit not-statistically significant) sign on the BirdPct covariate. This is the scope covariate, and would suggest a negative relationship between program size and willingness to pay.

In contrast, the MBP model has a much better Wald statistic. More importantly, the sign on BirdPct is both positive and significant. Considering WTP, all 4 MBP models show approximately a 25% increase in WTP when program size is increased. Not unexpectedly, the "participants only" variants show larger absolute values (since non-participants would tend to have lower E[WTP] values).

The two-stage models provide a more natural means of using both the participation and the response (given participation) questions. By allowing different factors at each stage, or by allowing similar factors to have differential impacts at each stage, the use of a two stage model may more accurately model how respondents consider their choices.

The results in Table 1b and 2b are mixed. The single bounded models still perform poorly (though BirdPct has a positive, albeit insignificant, impact on the probability of paying something). The multiple bounded models still perform fairly well. However, although BirdPct is significant and positive in both stages of the two-stage multiple bounded models, the net effect on WTP is small.

Of greater interest is the correlation coefficient. For the single bounded model, it is large and negative, whereas it hits the +1.0 constraint in the multiple bounded model. Either of these results are somewhat hard to explain, and suggest possible pathologies in the estimator.

To explore the estimator quality, two variants of a simulated dataset were created. The first uses fully synthetic data and coefficients, while the latter uses actual survey data and MBP estimates of coefficients. In both cases, a bivariate normal error term (with known covariance matrix) is generated and added to the known value of μ_i ($i=1,2$, $\mu_i = X_i \beta_i$ where both X_i and β_i are known). Participation occurs if $\mu_1 > 0$, and bounds are chosen to bracket μ_2 (from a set of "bids" ranging from 1 to 100).[7]

Table 3 lists the results of several simulations, one with relatively low participation rates, and one with relatively high participation rates . Examination of these results suggests that the MBP2 (two-stage MBP) model perform well, especially in terms of accuracy of predicting WTP. The MBP spike model does fairly well in predicting WTP, but can yield large errors on the constant, presumably due to inclusion of non-participants who may be systematically different (in terms of their first stage behavior) then participants. The participants only models (both MBP and SBP) perform similarly, they do a better job when the participation rate is higher.

However, it is striking how poor the coefficient predictions are for all the "high participation" rate models. This suggests that some caution may be needed when interpreting coefficient values.

Conclusions

Despite somewhat equivocal results, it appears that the use of multiple bounded probit models, especially in combination with information on "participation", can improve estimator efficiency. In particular, the two-stage multiple bounded probit and the simpler "spike" multiple bounded probit do well in the simulations, and yield better quality (lower standard errors) when applied to the Iowa data.

In addition to providing more accurate WTP estimates, these MBP models yields tighter confidence intervals around the "scope" variable. That is, where the single bounded model suggests that scope effects are insignificant (or even negative), the MBP models suggest that they are significantly positive.

---

[7] For single bounded simulations, a randomly chosen bid value is compared to μ_2 .

Table 3a: Simulated results (low participation scenario)

		Mean[X] (sd[X])	True β	Average β_error_ratio				
				MBP2	MBP, spike	MBP, partic	SBP, partic	SBP2
First Stage	CONST		2.0	0.05				0.05
	XA_1	0.77 (0.42)	-1.86	0.04				0.04
	XB_1	3.5 (1.9)	-0.55	0.04				0.05
Second Stage	CONST		1.2	2.6	49.1	9.0	6.7	3,.6
	XA_2	1.8 (1.1)	12.9	0.08	0.64	0.08	0.16	0.15
	XB_2	17100 (10100)	3.5e-4	0.26	0.71	0.30	0.48	0.47
	σ		22	0.07	1.9	0.09	0.11	0.11
	ρ		0.52	0.10				0.22
Average WTP_ratio				0.01	0.043	0.24[b]	0.21[b]	0.05

Table 3b: Simulation results (high participation scenario)

		Mean[X] (sd[X])	True β	Average β_error_ratio				
				MBP2	MBP, spike	MBP, partic	SBP, partic	SBP2
First Stage	CONST		-0.36	1.9				1.6
	X1 (birdpct)	54.2 (60.4)	0.0023	1.04				1.04
	X2 (income)	42957 (29901)	6.4e-5	0.82				0.74
	X3 (huntvu)	0/1, with 67%=1	0.444	0.96				1.04
Second Stage	CONST		-7.74	7.7	5.4	10.4	10.2	11.7
	X1	as above	0.047	1.2	1.3	1.1	1.2	1.2
	X2		1.3e-4	1.6	1.9	1.1	1.1	2.1
	X3		9.11	0.72	0.82	0.61	0.84	0.80
	σ		21	2.1	2.6	1.3	1.2	1.2
	ρ		0.2	3.3				4.4
Average WTP_ratio				0.40	0.45	0.74[b]	0.70[b]	0.64

Notes on Tables 3a and 3b:

1) For both scenarios, 50 iterations were used, with each iteration using a different synthetic datasets All dataset were constructed from the same "true" β and same covariates.

ii) Each dataset in the "low participation" scenario has 1300 "respondents". On average, 292 respondents were "participants" (as defined below).

iii) Each dataset in the "low participation" scenario has approximately 750 (variation in number of respondents is due to removal of internally inconsistent outcomes).. On average, 600 respondents were participants.

iv) Estimated models:
MBP2: 2-stage multiple bounded probit
MBP, spike: Multiple bounded probit, with "non-participants" assigned bounds of -infinity and 0.
MBP, partic: Multiple bounded probit, using participants only
SBP, partic: Single bounded probit, using participants only
SBP2: 2-stage single bounded probit

22) The average true WTP for all observations (for participants only):
Low participation: 9.00 (32.00)
High participation scenario: 35.00 (45.00)
Note that the "SBP, partic" and the "MBP, partic" models use participants only, hence average WTP will be larger.

vi) The average true wtp (WTP_true) is defined using:
$WTP\_true = \mu_2 + \varepsilon_2$, given $\mu_2 + \varepsilon_2 > 0$ (and, in "all observations"models, $\mu_1 + \varepsilon_1 > 0$))
$WTP\_true = 0$, otherwise
The average true WTP is taken across all iterations.

vii) $\beta\_error\_ratio = [\Sigma_{i=1..50}\ abs(\beta - b_i)/abs(\beta)\ ] / 50$
where β is the "known" coefficient value, and b_i is predicted coefficient from iteration i.

viii) ΔWTP_ratio_i (for iteration i) is defined as:
$$\Delta WTP\_ratio_i = [Avg_i(E[WTP_{ij}]) - Avg_i(WTP\_true_{ij})]/Avg_i(WTP\_true_{ij})$$
where:
E[WTP] is a "respondent's" expected willingness to pay, as computed using the predicted coefficients
$Avg_i(V_{ij}) = [\Sigma_{j=1..1300} V_{ij}] / 1300$; for some quantity V_{ij} (jth respondent in the ith iteration)

References

Alberini, Anna, Kevin Boyle, and Michael Welsh (2001). "Analysis of Contingent Valuation Data With Multiple Bids and Response Options Allowing Respondents to Express Uncertainty", under review at JEEM.

Arrow, K. R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H Schuman (1993). "Report of the NOAA Panel on Contingent Valuation". *Federal Register,* 58:10, 4601-4614

Bishop, Richard and Thomas Heberlein (1979). "Measuring Values of Extra Market Goods:Are Indirect Measures Biased". *American Journal of Agricultural Economics,* 61:5, 94-98

Boyle, Kevin J., Richard C. Bishop, Mary C. Ahearn, Daniel Hellerstein, Michael P. Welsh, Andrew Laughland, and John Charbonneau. 2000. "Test of Scope in Contingent Valuation Studies: Are the Numbers for the Birds?" Paper presented at the ASSA meetings, Boston, MA, January 6-9, 2000.

Carson, Richard T. 1997. "Contingent Valuation Surveys and Tests of Insensitivity to Scope." In R.J. Kopp, W.W. Pommerehne, and N. Schwarz (eds.), *Determining the Value of Non-marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Kluwer Academic Publishers: Boston.

Diamond, Peter A., and Jerry A. Hausman. 1993. "On Contingent Valuation Measurement of Nonuse Values." In, J.A. Hausman (ed.), *Contingent Valuation: A Critical Assessment*, pp. 3-38. North Holland: Amsterdam.

Greene, William (2000). "Econometric Analysis, 4th edition". Prentice Hall, Upper Saddle River, NJ.

Haneman, Michael, and Barbara Kanninen (1998). "The Statistical Analysis of Discrete Response CV Data". Working Paper No. 798, California Agricultural Experiment Station, 1-109

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

Haneman, 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

Kristom, Bengt. "Spike Models in Contingent Valuation". *American Journal of Agricultural Economics,* 79:3, 1013-1023

Maddala, G.S (1983). "Limited-dependent and qualitative variables in econometrics". Cambridge University Press, Cambridge.

Welsh, Michael P. and Gregory L. Poe. 1998. "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." *Journal of Environmental Economics and Management* 36, 170-85.

Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation

Thomas A. Heberlein
University of Wisconsin-Madison

Matthew A. Wilson
University of Maryland

Richard C. Bishop and Nora C. Schaeffer
University of Wisconsin-Madison

Presented at the Annual Meeting of Western Regional Project W-133
Miami Florida
February 26-28, 2001

This research was supported by NSF Grant SBR 9525533. The field work was conducted from the Kemp Natural Resources Station, and the Trout Lake Station.

We believe that the best way to establish the validity of values estimated through the contingent valuation (CV) method is to compare them with values estimated in what are now called simulated markets (Bishop and Heberlein 1979). Simulated markets involve laboratory or, better yet, field experiments where results from actual transactions can be compared with CV values. But such markets are often difficult to set up. In the struggle to assess the validity values estimated by CV economists have turned to the "scope test" a simple statistical test that could be more generally applied than simulated markets. A CV survey that offers a consumer/respondent more of an environmental good or a higher level of environmental services should elicit greater economic value (Mitchell and Carson 1989). Thus, observing whether CV respondents are (a) sensitive or (b) *in*sensitive to differences in the scope is appealing for its intuitive, commonsense logic and theoretical simplicity.

A recent review by Carson (1997), however, showed that out of 22 studies where scope sensitivity was investigated, four failed to show scope outright while two others showed mixed results (see Appendix A). Studies that failed to find scope have been widely discussed in the literature and have sometimes been used to discredit CV more generally as a valuation methodology (Diamond and Hausman 1994).

Even though most studies show scope, however, even a few exceptions are troubling in the face of questions about theoretical validity(Bishop, Champ, Brown, and McCollum 1997). What would the theory of gravity be if one out of five rocks when dropped landed on the ceiling rather than the floor? Perhaps those that go up are not really rocks, but helium balloons cleverly disguised as rocks. In order to better understand the utility of the scope test we need to better understand the *conditions* that produce non-scope. The object of this study is to examine such conditions—to search for the helium--for four environmental goods using both aggregate and individual data as well as retrospective interviews.

Previous Failures to Find Scope Sensitivity

The idea of scope *in*sensitivity actually first originated with Kahneman (1986). To support his hypothesis, Kahneman (1986) presented a graph (p. 191) showing three demand curves derived from a telephone survey of Ontario residents. Each curve was interpreted as representing respondent's demand for one of three *nested* goods: fishing in the lakes of (1) All of Ontario—the 'whole', (2) the Haliburton region in Ontario—a 'part' of the 'whole', and (3) the Muskoka region in Ontario-also a 'part' of the whole (Kahneman 1986).[1] All respondents were asked their willingness to pay (WTP) in the form of a tax to maintain the quality of fishing in these three geographically distinct regions. The resulting graph shows three demand functions that are very similar in shape and magnitude. In other words, "people seem to be willing to pay almost as much to clean up one region or any other, and almost as much for any one region as for all Ontario together" (Kahneman 1986, p. 191).

In addition to reiterating the original findings from the Ontario experiment, Kahneman and Knetsch (1992) provided new empirical data that has since been widely cited as further evidence of scope *in*sensitivity (Diamond and Hausman 1994) All respondents were given the same information describing an inclusive package of public

[1] The terms "part" and "whole" were first used in this context by Mitchell and Carson (1989, p. 237), who clearly anticipated scope issues when they spoke of the possibility of "part-whole bias."

services including education, health, police protection, roads, and environmental services. WTP questions were then administered to three sub-samples. Respondents in one sample received three WTP questions starting with the most inclusive good--environmental services, a subset of that inclusive good--improved disaster preparedness, and ending with the most specific good--improved rescue equipment and trained personnel. Respondents in a second sample were only asked two WTP questions: the first question asked about WTP for improved disaster preparedness and a subsequent question asked about WTP for improved rescue equipment and trained personnel. Finally, respondents in a third sample were asked about their WTP to improve the availability of equipment and trained personnel for rescue operations alone. Because Kahneman and Knetsch (1992) observed that no statistically significant difference between mean WTP for the public goods, the authors concluded that respondents were insensitive to the inclusiveness of the public good being valued and that the magnitude of the good had no discernible effect on WTP.

Three additional studies commonly cited as supporting the claim of scope insensitivity were carried out by Desvouges *et al.* (1993), Diamond *et al.* (1993), and Schkade and Payne (1993) (Hausman 1993). First to test whether the contingent valuation method was sensitive to scope variations, Diamond et al. (1993) elicited WTP estimates to avoid a 1% annual commercial timber harvest in several different wilderness areas throughout the Western United States. Diamond et al. (1993) tested the hypothesis that WTP would vary by the *size* of the wilderness area being protected. Using the same split-sample design, the three areas were the Selway Bitteroot wilderness (1.3 million acres), the Bob Marshall (1.0 million acres), and Washakie (0.7 million acres). No significant difference was found between WTP estimates for the three areas. Several other treatments involving wilderness areas also failed to find scope sensitivity. The authors concluded that in general, "whatever CV surveys may be measuring, they are not measuring consumers' economic preferences over environmental amenities" (Diamond, et al. 1993, p. 61).

Desvouges et al. (1993) investigated the sensitivity of WTP to prevent (a) 2000, (b) 20,000, or (c) 200,000 birds from being killed in oil holding ponds in the Central Flyway. The CV survey involved a self-administered questionnaire conducted with respondents in Atlanta shopping malls. Respondents in three different sub-samples were asked their WTP to prevent the deaths of either 2000, 20,000, or 200,000 birds. The resulting means for the three treatments were \$80, \$78, and \$88, were not statistically different, leading the authors to conclude that, "WTP estimates of nonuse values do not satisfy simple validity and reliability requirements...current methods for estimating nonuse values are neither valid nor reliable for damage-assessment purposes" (Desvouges et al. 1993)p. 93). This study was replicated by by Schkade and Payne (1993).

Rethinking Scope Sensitivity Using Social Psychological Attitude-Behavior Theory

Differing conceptions of "value" have long been recognized and debated in both economics (Milgrom 1993; Aaron 1994)and social psychology (Heberlein, 1988; Peterson et al. 1988; Schwarz 1997). Traditionally, economic theory has defined 'value' in a rather strict and narrow behavioral sense(Freeman 1993; Milgrom 1993). In this sense, the economic value of a commodity is no more and no less, than the amount of money a person is willing to give up to get the commodity, or the amount the person requires as compensation for loss of the commodity. On the other hand, in terms of attitude theory, willingness-to-pay as an elicited contingent value is best conceived as a behavioral intention—an expression of a *willingness* on the part of the survey respondent to engage in a behavior relevant to the commodity(Ajzen and Peterson 1988, Heberlein 1988). A CV value is *not* an observable behavior like buying or selling commodities in the marketplace; rather it is an expressed intention to make a purchase should the opportunity be available.

Attitude theory suggests this behavioral intention is influenced by affective and cognitive dimensions of a person's "attitude" towards a commodity (Zajonc et al. 1982; Zajonc 1980). On the one hand, 'cognitions' involve the thoughts and knowledge that people might have about an environmental commodity. For example, a survey respondent might say, "I know a lot about air quality" or "I think a lot about the air that I breathe". Cognitions are generally conceptualized as *information, knowledge or beliefs*, where beliefs are understood to be the associations or linkages that people establish between the attitude object and various objective attributes(Fishbein and Ajzen 1975). The affective dimension of attitudes deals with emotions. An example might be a statement such as, "I like the air quality in my neighborhood."

Scope tests have traditionally looked at scope in the behavioral intention domain of attitudes. But attitude theory would suggest that we might also want to look at 'affective scope' -- liking the whole more than the part—or 'cognitive scope'--knowing more and thinking more about the whole than the part. Under these conditions, we might reasonably anticipate that the respondent will exhibit all of the virtues of a rational consumer and express a higher WTP (behavioral intention) for more of the commodity than less. Conversely, a respondent may show something akin to 'reverse affective scope'—liking the part more than the whole—or 'reverse cognitive scope'—knowing more and thinking more about the part than the whole. In this case, attitude-behavior theory suggests that a higher WTP (behavioral intention) will be expressed for the part of the environmental good rather than the whole. In this manner, scope *in*sensitivity is placed in a richer theoretical context; one that appreciates both the attitudinal and situational characteristics that might reasonably lead the individual to show scope *in*sensitivity.

Our point is to move beyond merely accepting or rejecting the validity of estimates derived using the contingent valuation method, and instead describe the specific factors that affect scope judgments in a real world context.

Four Attitude Objects (Environmental Goods)

Based on 27 interviews with randomly selected property owners in Vilas and Oneida counties in Northern Wisconsin (the "Lakeland Area), we identified four environmental goods for study. **Water quality** in lakes was selected because it is a concrete and symbolic object of vital economic and social importance to residents of the Lakeland Area. The part was the well-known Minocqua chain of lakes in the center of the study area. The whole was all of the lakes in Vilas and Oneida County (over 2300 lakes in all) including the Minocqua chain.

The second object was wolves, or more precisely, **wolf populations** in Northern Wisconsin. The current population is about 200 wolves (Thiel 1993). At the time of our research, the Wisconsin Department of Natural Resources was developing a wolf management plan and the question of wolf populations in northern Wisconsin was being publicly debated. Wolf populations were numerically nested: the whole was 800 wolves in Northern Wisconsin and the part was 300. Wolves are symbolic of wildness in nature, but unlike water quality in lakes, most people don't get to see wolves or otherwise interact with wolves.

Policy makers would sometimes like to have economic values for complex scientific concepts. Because we are skeptical of the adequacy of CV for valuing such domains, which do not have the concreteness of lakes or even wolves, we wanted to include such a good. In our developmental interviews we asked about **biodiversity** Most Lakeland property owners thought of biodiversity as a simple proxy for the concept of "nature". For example, while deer in northern Wisconsin are overabundant and biologists complain that their presence actually reduces the number and distribution of plant species, most residents we interviewed felt that deer *added* to biodiversity. Biodiversity is ambiguous and confusing but important. The whole was protection biodiversity in all of Northern Wisconsin, while the part was protecting biodiversity only in Vilas and Oneida counties.

In the 1980's local Indians had won state and federal court cases that reestablished their right to hunt, fish and gather off reservation. The Indians began to exercise their rights to harvest game fish with spears during the spawning season in April, before the regular sport fishing season opened. This created controversy and confrontations at boat landings that bordered on race riots. Although the research team went into the field thinking that the conflict was long over, the topic came up repeatedly in our developmental interviews with respondents expressing very strong feelings about the issue. In an effort to have one environmental object where perhaps rationality was overwhelmed by emotion, we selected **Chippewa Indian spear fishing** in the Minocqua Chain (part) and in all of the lakes in Vilas and Onieda county (whole).

Mail and Telephone Survey

Information about knowledge, interest, and satisfaction with the whole and the parts of the four attitude objects was obtained using a 19 page mailed questionnaire. The inside cover had a map of the two counties where respondents could circle the lakes they had experience with and a color map showing the state with northern Wisconsin and Vilas and Oneida Counties highlighted so people could see what we meant by the "whole and the part areas."

One week after we received the completed questionnaire, respondents were contacted by phone to measure their willingness to pay The first telephone interview asked respondents what they would be willing to pay for all four objects, two parts and two wholes. The question-order sequence of objects was randomly assigned. Furthermore, during the first telephone interviews (referred to below as the Time 1 interview), each respondent was randomly assigned either the part or the whole for each item. Two weeks after their first completed telephone interview, respondents were contacted for a second interview (the Time 2 interview), which obtained their willingness to pay for the remaining four part-whole complements.

Procedures and Response Rate

The final sample size for the mail questionnaire was 1,435 cases (an additional 65 questionnaires were undeliverable or the individual was deceased or no longer living at the specified address). One hundred and twenty four respondents either refused participation by returning their mail questionnaire with a note stating that they did not want to participate in the study or by telling the interviewer that they did not wish to continue with the study during the reminder telephone call conducted on November 6, 1998. By January 25, 1999 the final number of completed mail surveys was measured at n=876, with an overall mail response rate of 876/1,435 = 61%.

For the telephone interviews, response rates and sample dispositions were calculated both separately and cumulatively for Time 1 interviews and Time 2 interviews. As anticipated, attrition rates and observed non response errors were substantially higher for Time 1 telephone interviews than Time 2 interviews, with 70 cases refusing outright to participate in the first telephone interview and 120 determined to be "not available" by the UW survey center staff after repeated calls. The final within-mode response rate for telephone Time 1 was 686/876 = 78% with a cumulative response rate of 686/1,435 = 48%. In the case of Time 2 interviews, attrition rates and non-response declined to only n=29 refusals with only n=49 cases determined to be not available. The final within-wave response rate for Time 2 was therefore 617/676 = 90% with a cumulative response rate of 617/1,435 = 43%. Hereafter, the final sample of n=617 cases described above will serve as the primary data used in this study and will be referred to as the Lakeland survey sample.

Independent Variables: Affect, Cognition, Experience and Personal Characteristics

In the initial three pages of the mailed survey respondents were asked to report where they lived in the Lakeland Area, if it was a seasonal residence, and how much time they spent there. They were also asked to report their participation in outdoor sports and in local environmental activities. The questionnaire also included a 6-item modified New Environmental Paradigm (NEP) Scale (Dunlap and Van Lier 1984; Dunlap and Van Liere 1978). The next 14 pages measured attitudes toward the four objects (part and whole). The last 3 pages measured respondent's attitudes toward payment and standard social and economic variables.

Table 1: Independent Variables Used in Analysis

Variable	Example Description
NEP	Attitude scale representing ecocentric value-orientation
Importance	How important is the issue of lake water quality to you personally? 1 = Not at all important..............7 = Extremely Important
Feelings	How do you feel about the quality of water in Northern Wisconsin? 1 = Strongly Dislike..................5 = Strongly like
Think About	How often do you think about the quality of water in Minocqua Chain (All Lakes)? 1 = Never...............................6 = Almost Every Day
Know About	How much would you say you know about the quality in Minocqua Chain (All Lakes)? 1 = Know Almost Nothing......... .5 = I'm An Expert
Satisfaction	How satisfied are you with the quality of water in Minocqua Chain (All Lakes)? 1 = Extremely Dissatisfied...........5 = Extremely Satisfied
Affect	How would you rate the quality of water in the Minocqua Chain (All Lakes)? 1 = Extremely Bad..................... 7 = Extremely Good
Recreation Experience	Experiences associated with Minocqua Chain of lakes (All Lakes): 1= No experience......................5 = A great deal of experience
Affect Strength	Cumulative Affect (Satisfaction & Affect) towards Minocqua Chain and All Lakes 0 = No feelings......................24 = A lot of Feeling
Cognition Strength	Cumulative Cognitions (Think & Know) towards Minocqua Chain and All Lakes 0 = No Cognitions.....................22 = A lot of Cognitions
Affect Scope	Affect (Satisfaction & Affect) Difference between Minocqua Chain and All Lakes -12............ 0= No Difference........+12
Cognition Scope	Cognitive (Think & Know) Difference between Minocqua Chain and All Lakes -11............ 0= No Difference........+11
WI Residence	"Did your household pay Wisconsin State Income Tax in 1997?"
Angler	"I spend time fishing on lakes in Vilas and Oneida counties and think of myself as a committed angler"
Age	"In what year were you born?"
Gender	"Are you?" 0 = Male 1 = Female
1997 Income	"What was the combined total 1997 annual income, before taxes of all members of your immediate family living in your household?"

For all environmental objects, the brief introduction was followed by two broad attitudinal questions—one question asked respondents how important water quality, wolves, spearing, and biodiversity was to them personally and a second asked how they felt about the object (strongly dislike to strongly like). Following the two questions, the the differences between the environmental whole and the environmental part were described in a special box (e.g., "Vilas and Oneida counties includes only the area that lies within the two-county boundary" and "All of Northern Wisconsin includes all of the state of Wisconsin north of Highway 8).

Following the part-whole distinction, the questionnaire proceeded to measure four attitudinal variables for **both** the whole and the part. First, respondents were asked to rate each object on a modified Likert scale ranging from extremely bad to extremely good. Second, respondents were asked to report how much they knew about the environmental object on a five-point scale-- "I know almost nothing" to "I am an expert". Third, respondents were asked to rate their personal level of satisfaction with the current state of the environmental object on a five-point scale ranging from "extremely dissatisfied" to "extremely satisfied". Finally, respondents were asked to report how often they thought about the object in question on a six-point scale from "never" to "almost every day".

In the case of lakes, we were able to measure respondent's direct experience with lakes with a five-point scale of participation on both the Minocqua Chain and all lakes of Vilas and Oneida counties (i.e. whether the respondent had power boated or water skied, fished, canoed or sailed, swum, or participated in shore line activities).

Dependent Variables: Contingent Values and Scope

Although the senior members of our research team pioneered the dichotomous choice format for CV in the 1970's, we chose to use an open ended approach in the current study. One problem with dichotomous choice is loss of information. If a person will pay 50 dollars in a take it or leave it format, we have no idea if he or she would really pay $150. The open-ended format allowed us to gain a point estimate of WTP at the individual level. This, in turn, allowed us to conduct scope tests at the individual, as well as the aggregate, level.

Our developmental interviews showed that many respondents were averse to conventional CV payment vehicles like property taxes when it came to issues like reducing the amount of spear fishing in lakes, yet were amenable to the idea of using taxes to increase the number of wolves in the area. On the other hand, at various points in our developmental interviews, respondents spontaneously brought up the idea of voluntary contributions to protect and increase biodiversity or mentioned things like higher construction and building permit fees to protect water quality in the lakes. A one-time payment to an "environmental trust fund" was selected as the payment strategy (Stevens et al. 1991; Spash and Hanley 1995). We offered respondents a range of four payment alternatives that would be directed into the public trust fund: (1) a one-time voluntary donation; (2) a one-time levy on property taxes, (3) a one-time levy on state income taxes or (4) one-time charges or fees for things like new housing and construction permits.

Box 1: Sample Willingness to Pay Question

One way to raise money would be for people to pay into *a public trust fund* that would be set aside by the State of Wisconsin to increase the number of wolves from 200 to 800...

...At this time, we don't know how you might be asked to pay into the trust fund by the State of Wisconsin, but we do know that payments to the fund would take place on a one-time basis and money could be collected in one of the following four ways:

- You might pay directly to the trust fund through a one-time voluntary donation.
- Your property taxes might increase on a one-time basis, affecting you directly through your tax bill or indirectly through the rent on your residence.
- If you are a Wisconsin resident, your state income taxes might increase on a one-time basis
- Or, you may pay directly through one-time government charges and fees on things like new housing construction, well drilling, septic system, and other permits.

Now, suppose that the number of wolves in Wisconsin could be increased from 200 to 800 if enough money were raised by the Trust Fund.

If you were given the one-time opportunity to pay money to the Trust Fund, what is the most money you would be willing to pay to ensure that the number of wolves in Wisconsin is increased to 800?

Findings

Attitudes Toward the Objects

The data revealed four environmental goods that were viewed quite differently by respondents. Water quality, spear fishing and biodiversity were all important but our respondents strongly disliked spearfishing while having positive feelings toward water quality and biodiversity. Wolves were seen as not important or unimportant by almost a majority of the respondents and nearly a majority were either neutral or disliked wolves. (Table 2)

Table 2: Percent Rating Importance and Feelings Toward Four Environmental Goods

Variable		Water Quality	Wolves	Spearfishing	Biodiversity
Importance					
	Not Important	0.3	11.2	2.5	1.2
	Very Unimportant	2.3	4.9	1.6	2.8
	Somewhat Unimportant	0.2	5.6	2.3	3.6
	Neither	0.2	26.1	7.7	7.2
	Somewhat Important	9.0	31.9	24.3	27.2
	Very Important	47.5	14.1	28.4	41.5
	Extremely Important	40.5	6.2	33.2	16.5
		100	100	100	100
	Mean	6.2	4.3	5.7	5.5
	S.D.	.96	1.6	1.4	1.2
Feelings					
	Strongly dislike	0.7	4.7	49.9	0.2
	Dislike	8.9	6.5	32.8	4.8
	Neutral	14.3	33.9	14.1	25.2
	Like	60.5	41.7	2.1	60.6
	Strongly like	15.7	13.1	1.0	9.2
		100	100	100	100
	Mean	3.82	3.52	1.7	3.7
	S.D.	.83	.96	.86	.69

Aggregate Scope

Using only Time 1 interviews, we were able to test for scope in mean values in a the split sample design. WTP to maintain water quality and prevent spear fishing showed aggregate scope. (Table 3). These same goods also showed cognitive, affective and direct experience scope. Respondents tended to say that they knew more about and thought more about water quality and spear fishing in all lakes than in the chain, and that they liked the water quality and disliked spearing more in all the lakes than in the chain. They also recreated more in all the lakes than the chain, our measure of direct experience scope.

Wolves failed to show aggregate scope. People on average were no more willing to pay for 800 than they were for 300. They failed to show cognitive or affective scope as well. Actually they showed reverse scope in their attitudes. Our respondents knew more and thought more about 300 wolves than 800 and they were much more satisfied with 300 wolves than 800.

In the aggregate respondents were significantly more likely to pay to maintain biodiversity in their local area than they were in the whole North (i.e., reverse aggregate scope for WTP). They also showed reverse cognitive and affective scope.

Table 3: WTP and Attitude Scope for Four Environmental Goods

	Part	Whole		
Water Quality	Minocqua Chain	All Lakes	**Difference**	**Scope**
Willingness to Pay	$107	$260	**+$153**	Yes
Know about	2.23	2.59	**+ 0.36**	Yes
Think about	2.82	3.44	**+1.62**	Yes
Satisfaction	3.42	3.73	**+0.31**	Yes
Affect	4.97	5.50	**+0.33**	Yes
Recreation Experience	2.50	4.23	**+1.73**	Yes
Spear fishing	Minocqua Chain	All Lakes		
Willingness to Pay	$47	$102	**+$55**	Yes
Know about	3.23	3.47	**+0.24**	Yes
Think about	2.84	2.95	**+0.11**	Yes
Satisfaction	2.25	2.13	**-0.12**	Yes*
Affect	2.26	2.13	**-0.13**	Yes*
Recreation Experience	2.50	4.23	**+1.73**	Yes
Wolves	300 Wolves	800 Wolves		
Willingness to Pay	$42	$40	-$2 ns	No
Know about	1.97	1.71	**-0.26**	No
Think about	2.34	1.94	**-0.40**	No
Satisfaction	4.44	3.48	**-0.96**	No
Affect	4.49	3.46	**-1.03**	No
Biodiversity	2 Counties	N. Wisconsin		
Willingness to Pay	$173	$125	**-$48**	No
Know about	2.42	2.29	**-0.13**	No
Think about	2.96	2.79	**-0.17**	No
Satisfaction	5.09	5.03	**-0.06**	No
Affect	5.11	5.04	**-0.07**	No

Values in bold represent p > .05

*The affect and satisfaction questions measure how satisfied the respondent is with the "current level" of off reservation spearfishing by Chippewa. The CV question asks the respondent how much money he or she would be willing to give to halt the current level of off reservation spearfishing. Affective scope implies that the respondent is willing to pay more to reduce spearfishing where they are more dissatisfied with the current level. Thus, the *negative* difference score respresents positive affective scope.

Individual Scope

Studies cited previously that failed to show scope were based on split sample designs. That is, people got to express a WTP for only a whole or a part but not both.

Thus the only analysis that could be done was to look at averages as we did in the last section. But human behavior is complex and much is hidden behind averages. By asking people for their WTP for both the whole and the part in two telephone interviews separated by a week allows us to explore scope at the individual level. This allows us to reconceptualize non scope: 1) a person's WTP can be zero for the whole and zero for the part, we call this "zero no scope," 2) a person's WTP can be positive but the same amount for both the whole and the part, 3) a person's WTP can be less for the whole than the part, something we might call "reverse scope." (Table 4)

Table 4: Scope and No Scope for Aggregate and Individual Data

AGGREGATE SCOPE	NO SCOPE $WTP_P \geq WTP_W$			SCOPE $WTP_P < WTP_W$
INDIVIDUAL SCOPE	REVERSE SCOPE $WTP_P > WTP_W$	ZERO NO SCOPE $WTP_{P=0} = WTP_{W=0}$	POSITIVE NO SCOPE $WTP_{P>0} = WTP_{W>0}$	POSITIVE SCOPE $WTP_P < WTP_W$

The results of the individual analysis are presented below in Figure 3. As the data show, *a majority of respondents failed to show scope for each of the four environmental objects*. For **water quality**, which showed scope in the aggregate, only 41% of respondents showed scope sensitivity—that is, only 2 out of 5 respondents were willing to pay more to clean up all of the lakes than the four lakes in the Minocqua chain. For Indian spear fishing which also slowed aggregate scope only 1 in 5 showed scope.

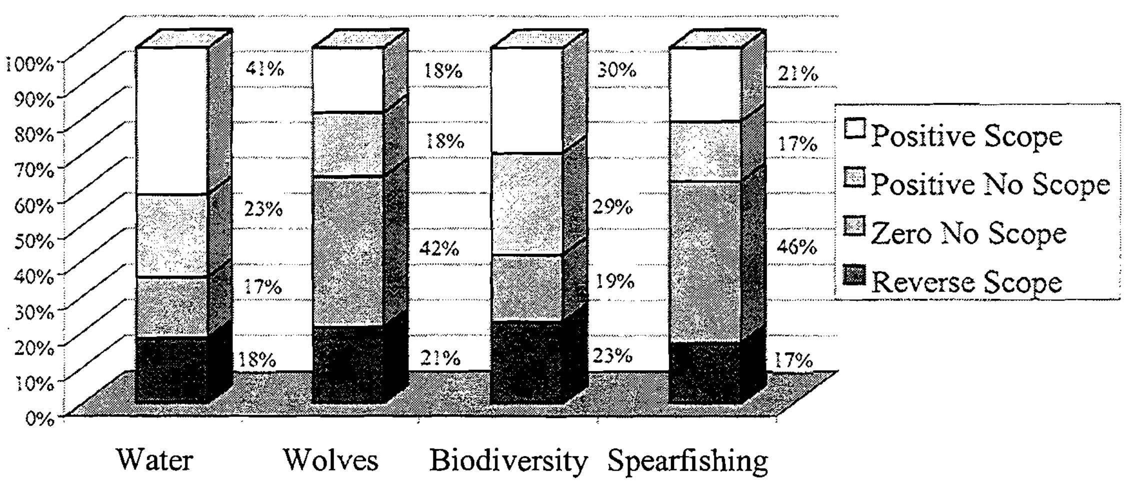


Figure 3: Four Types of Scope Sensitivity

For wolves, about which most people had little interest and no strong feelings, 42% said they would pay nothing for 300 or 800 wolves. The public thought spear fishing was important and had strong feelings about it but 46 percent said they would pay nothing to stop spear fishing on either the chain or all of the lakes. Biodiversity, which showed reverse scope in the aggregate, only had 23 percent of the people showing reverse scope when we consider the individual data.

Predicting Individual Scope Types

Because we have WTP for the whole and the part for each individual we can explore what attitudinal and personal characteristics explain which type of scope a person showed. There are four dichotomous dependent variables in this analysis: 1) positive scope compared to the other three groups, 2) positive no scope compared to the other three groups, 3) zero scope compared to the other three groups and 4) reverse scope. The independent variables include a general measure of environmentalism, belief that water quality is important to the respondent, general liking water quality in the north, affective scope(liking the whole more than the part), cognitive scope (knowing and thinking more about the whole than the part), cognitive strength(knowing and thinking about the good), affective strength (liking the good), recreation scope(using the whole more than the part (lakes only)) and five personal characteristics, age, gender, income, residence, and angling behavior.

If we can explain some of the variance in the scope types from these variables using logistic regression we have a better understanding of why people are expressing what seem to be inconsistent economic preferences. If such predictions prove impossible, then we might conclude that either random error or other unmeasured variables determined relative WTP.

Water Quality

The independent variables significantly predicted each of the four scope types (Table 5) for water quality.

Table 5: Logistic Regression Estimates for Water Quality Scope Sensitivity

Independent Variables	Positive Scope	No Scope Positive	No Scope Zero	Reverse Scope
Environmentalism	1.07	**1.14***	**0.82**	0.90
Water Quality Important to Respondent	0.96	1.05	1.23	0.89
Like Water Quality in Northern Wisconsin	1.37	.81	.61	1.10
Affective Scope	1.09	.94	.89	.97
Cognitive Scope	**1.14**	1.01	.96	**.81**
Cognitive Strength	1.03	1.00	.95	.96
Affective Strength	**.89**	1.05	1.11	1.07
Recreation Scope	**2.31**	**.37**	1.56	.71
Angler	1.14	.76	1.29	.94
Wisconsin Resident	1.30	**.50**	1.53	1.15
Age	1.00	.85	**1.38**	.90
Female	.85	.94	.51	**2.09**
Income	**1.13**	.96	**.91**	.94
Cox and Snell R Square	**0.13**	**0.08**	**0.11**	**0.07**
Model X^2 p value	p<.000	p<.011	p<.002	p<.000

*Values in bold represent Wald Statistic with p <.05. Coefficients over one increased odds of being in the category compared to all other categories and those below 1.0 decrease the odds. A coefficient of two means that a unit increase in the independent variable doubles the odds while a coefficient of .50 means that the person is half as likely to be in the category.

People who show cognitive scope are more likely to be in the positive scope category. Respondents who show recreation scope are more likely to show positive scope. If you feel the water quality is better in both the whole and the part (affective strength) then you also are willing to pay less to maintain quality. Those with higher incomes show positive scope.

Respondents who are willing to pay some positive amount but the same for the whole and the part (**positive no scope**) are less likely to be Wisconsin residents, much less likely to have recreational experience on the area lakes and more likely to hold pro environmental attitudes.

Those who are willing to pay zero for both the part and the whole **(zero no scope)** are older, have lower incomes, and are less likely to hold pro environmental values. They give lower ratings to the quality of the water in lakes all over the north.

Those who show **reverse scope**—who are willing to pay more to maintain water quality on the chain and are less likely to show cognitive scope--say they know and think more about water quality on the Minocqua chain than the rest of the lakes. They are also more likely to be female.

A number of variables had nothing to do with any of the water quality scope types. Anglers were no more or less likely to fall into any of the categories. Neither those showing affective scope (liking the whole more than the part didn't have any effect) nor those who thought water quality was important show any differences. Cognitive strength (thinking a lot and knowing a lot about water quality) didn't help explain scope. Those who liked the water quality in the north or for whom water quality was important were also no more likely be in one of the scope categories. Environmental attitudes did differentiate the types but in different directions. Respondents who showed positive no scope held pro environmental attitudes, while those who showed zero no scope were less likely to do so.

Spear Fishing

When we turn to spear fishing scope, the most notable thing is that the same independent variables that helped explain water quality scope in the same set of lakes were not able to explain the scope types for spear fishing. Only one of the four logistic regression models was significantly different from zero. (Table 6)

Table 6: Logistic Regression Estimates for Spear Fishing Scope Sensitivity

Independent Variables	Positive Scope	No Scope Positive	No Scope Zero	Reverse Scope
Environmentalism	1.02	0.96	1.00	.99
Spear Fishing Important to Respondent	1.06	.97	.98	.94
Like Spear fishing in Northern Wisconsin	.84	**.43***	**1.88**	.90
Affective Scope	.92	1.34	1.08	.82
Cognitive Scope	1.06	1.09	.93	.95
Cognitive Strength	1.06	1.00	.95	.98
Affective Strength	1.06	1.11	**.88**	1.02
Recreation Scope Lakes	1.18	1.09	1.08	.67
Angler	1.34	1.06	**.65**	1.27
Wisconsin Resident	.72	.69	**2.16**	.73
Age	.95	.90	1.02	1.13
Female	**1.99**	.75	**.57**	1.15
Income	1.01	.95	1.00	1.04
Cox and Snell R Square	0.04	0.04	**0.07**	0.02
Model X^2 p value	NS	NS	p<.007	NS

*Values in bold represent Wald Statistic with p <.05. Coefficients over one increased odds of being in the category compared to all other categories and those below 1.0 decrease the odds. A coefficient of two means that a unit increase in the independent variable doubles the odds while a coefficient of .50 means that the person is half as likely to be in the category.

The only scope type that was significantly predictable was **zero no scope**—willing to pay nothing to prevent Indians from spear fishing on either the chain or all of the lakes. Anglers who are competing with Indians for fish were less likely to pay zero, but anglers were no more likely to show positive scope or reverse scope. Out-of-state residents were more likely to pay something to stop spear fishing, but this variable explained none of the other scope types. Men showed a greater likelihood of zero scope than women. Those people who liked spear fishing more (or who disliked it less) were much less willing to pay anything to stop it.

Wolves

Although wolves did not show aggregate scope, it turns out using individual data we can predict who will pay nothing for 300 OR 800 wolves. The R square predicting the **zero no scope** category was .38 (Table 7). The next highest R square in the other 15 regression models was .13. Wisconsin men with less positive environmental attitudes, who think wolves are unimportant and are less satisfied with wolves are the most likely to say they would pay zero dollars to preserve either 300 or 800 wolves. Or conversely out of state pro-environmental females who think wolves are important and feel positively about wolves are the least likely to be in the zero no scope category.

Table 7: Logistic Regression Estimates for Wolves Scope Sensitivity

Independent Variables	Positive Scope	No Scope Positive	No Scope Zero	Reverse Scope
Environmentalism	1.07	1.10	**.83***	1.04
Wolves Important to Respondent	1.02	**1.34**	**.67**	1.26
Like Wolves in Northern Wisconsin	.98	1.09	.66	1.36
Affect Scope	1.02	1.10	.92	.94
Cognitive Scope	1.23	1.02	.93	.88
Cognitive Strength	.97	.96	1.06	**.92**
Affective Strength	**1.21**	1.06	**.83**	.98
Wisconsin Resident	**.55**	.77	**3.19**	.86
Age	1.12	**.80**	1.17	.94
Female	1.68	**1.70**	**.39**	.92
Income	**1.07**	.96	1.04	.95
Cox and Snell R Square	**0.11**	**0.12**	**0.38**	**0.07**
Model X^2 p value	P<.000	P<.000	P<.000	P<001

*Values in bold represent Wald Statistic with p <.05. Coefficients over one increased odds of being in the category compared to all other categories and those below 1.0 decrease the odds. A coefficient of two means that a unit increase in the independent variable doubles the odds while a coefficient of .50 means that the person is half as likely to be in the category.

Those who show positive scope, who are willing to pay more for 800 than 300, like 800 more, are generally more satisfied with wolves and are more likely to be older females who live outside of Wisconsin. They also are likely to have higher incomes.

Respondents who give a positive value but don't differentiate between the whole and the part **(positive no scope)** are older people, females, and people who say wolves are important to them. **Reverse scope** for wolves is the least well explained by the independent variables. Those who say they have greater knowledge and think more about wolves are less likely to show reverse scope.

Biodiversity

The biodiversity story is much simpler to tell (Table 8). As these results show, none of our explanatory variables predicts **positive scope** or **reverse scope**. The logistic regression can explain who will pay nothing **(zero no scope)** for biodiversity in either the

Lakeland Area or all of northern Wisconsin and who will pay the same amount for each. It is your environmental attitudes. Going up one unit on the attitude scale increases the likelihood that you will be in the **positive no scope** and going up one unit decreases the likely hood you will pay nothing for biodiversity anywhere.

Table 8: Logistic Regression Estimates for Biodiversity Scope Sensitivity

Independent Variables	Positive Scope	No Scope Positive	No Scope Zero	Reverse Scope
Environmentalism	1.00	**1.11***	**.79**	1.07
Biodiversity Important to Respondent	.93	.97	.94	1.21
Like Biodiversity in Northern Wisconsin	.95	1.25	1.05	.79
Affective Scope	.94	.99	1.08	1.01
Cognitive Scope	.78	1.07	1.24	1.08
Cognitive Strength	1.02	1.02	.99	.95
Affect Strength	.95	1.05	1.02	1.00
Wisconsin Resident	.90	.89	1.60	.94
Age	.99	.88	1.22	.99
Female	1.34	1.23	.54	.78
Income	1.03	1.00	.93	1.01
Cox and Snell R Square	0.02	**0.05**	**0.11**	0.03
Model X^2 p value	NS	P<02	P<000	NS

*Values in bold represent Wald Statistic with p <.05. Coefficients over one increased odds of being in the category compared to all other categories and those below 1.0 decrease the odds. A coefficient of two means that a unit increase in the independent variable doubles the odds while a coefficient of .50 means that the person is half as likely to be in the category.

Post Survey Interviews

The research team conducted 30 retrospective interviews with respondents who had completed all three waves of the survey. Respondents were purposively sampled in order to ensure representation from all of the scope conditions (reverse scope, positive scope, zero no scope and positive no scope) across the 4 attitude objects. Respondent debriefing framed the purpose of the question-answer process about the survey, to gain a better understanding of how individual respondents interpreted the questions that we asked them (DeMaio and Rothgeb 1996; Willis et al. 1999). Retrospective think-aloud protocols were then used so that each respondent would "think aloud" as they read through their answers to the survey questionnaire, thus verbalizing the contents of their personal memories about why they answered each question in the manner that they did (Sudman et al. 1996).

Reverse Scope Is Not Irrational

The logistic regression analysis showed that at least some of the variance in this scope type can be explained by our independent variables. And our interviews showed why. After conducting several interviews the economist on our team observed, "These people aren't showing scope, but they aren't stupid." He was right. A respondent who was willing to pay $500 dollars for 300 wolves and $100 for 800 wolves observed: "I rated 300 wolves as very good and 800 wolves as good. And the difference is because of the potential for the very kind of public dissatisfaction that we're seeing right now…Oh yeah, it [800 wolves] would be a problem for the wolves and it would be a problem for the wolf advocates and so on…So therefore, I would um, I would be concerned about whether 800 wolves would cause problems"

This person was not the only one concerned about this. Another respondent who said they would pay $200 for 300 wolves and $100 for 800 wolves held the same beliefs. "I think 800 is a too many where it could become a problem for them and possibly a people problem with, or problems with us…. Thinking that 300 is probably closer to the reasonable number than 800, I would be more willing to support the 300."

People showed reverse scope on water quality for different reasons than they did for wolves. As a respondent who was willing to pay $500 to maintain water quality on the chain and $100 for all the lakes explained "I'm familiar with Lake Minocqua and to some extent Tomahawk Lake [lakes of the Minoqua Chain]…And, and I'm less familiar with other lakes in Vilas and Oneida county although I'm familiar with the large flowages, Rainbow Flowage." "I had a larger stake in the Minocqua chain and…so I figured I was more responsible as a, as a property owner for the Minocqua Chain of Lakes than for, um, Vilas County." The chain had a special meaning for the respondent and was not seen as a part of a larger whole. "The chain is where my house is." Another female claimed.

Zero Means Zero

Our regression analysis showed however that zero no scope was the most predictable. Indeed it was the only kind of scope that we could significantly predict across all four goods. In the case of wolves 11 variables explained almost 40 percent of the variance in the zero-zero dependent variable.

In the words of one respondent "I think we could get along without them [wolves] very well, which we have…I don't think they're very important, I don't think they're very unimportant. Nature kind of takes care of itself." Another told us, "I would rather see them do other things with tax money than increase the wolf population. 800 no, no, I don't think we should have 800 wolves anyway and I wouldn't pay penny one to get 800 wolves."

Even though people liked "biodiversity," 19 percent held views like this person: "I myself don't feel that I should pay to uh keep plants and animals you know…It's the people that are developing the property and ruining that is the ones that should pay for it. So if the developer wants to take out 40 acres and put a subdivision in, then there should be a tax or whatever you want to call it on his property and not the person that just is trying to eke out a living in his space, and that's basically my feelings"

When it comes to spear fishing, 46 percent of the respondents were willing to pay nothing to stop spear fishing even though they had strong negative feelings about it. Some don't think it will be effective or appropriate: "No, you won't buy those people off. (pause) You know this isn't the first thing on the spearing and fishing --probably say maybe 30 years ago they did come up with a reservation fishing license, and they have one now but it didn't go over at all… You're not going to change anything until you change people." Another gives us insight that the whole and the part are the same and he won't pay to stop spear fishing in either location "I didn't answer differently at all because there's no difference to me in where the spear fishing occurs. I didn't differentiate between the Minocqua chain and all the lakes in Vilas and Oneida county for any of the spear fishing questions. Because...the issue to me is identical regardless of where it happens. And then because I don't have, I don't feel a personal stake in any of these lakes...you know I haven't adopted any walleyes in any particular lakes or done any of the kind of wacky things that some fishermen might do. "

Positive Scope Isn't Always Well Informed

When one thinks about the no scope types we often reasons that there is just some kind of error. If only people were told that 800 wolves won't do any more damage or be any greater risk than 300 they might change their values. But in the case of wolves and biodiversity where people often had little information and very weak attitudes we got the sense the positive scope category was not without its own instability. This person based a response of $30 for 300 wolves and $75 for 800 on the following reasoning "I know very little. 800 wolves living in northern Wisconsin? Almost none. Hardly any information. The only information I have on the subject is what I've heard on the news. There's one group I believe up in the national forest over by (pause) I forget the name of the town, but there is one...Pack that they have, over the last 5 years I believe. I don't know if it's the Nicolet National Forest now, or if it's the other one... I believe it's Chequamegon....How satisfied are you or would you be with 300 wolves living in northern Wisconsin? Extremely satisfied. 800 wolves living in northern Wisconsin, here again extremely satisfied, and I would probably prefer the larger number of wolves...you'd have a better chance of seeing them, hearing them, observing them. That would be my, knowing that they are here." Notice that this positive scope was not based on a lot of information. If we had given him more information on the survey he might have ended up as a non scope type.

Positive No Scope Ignores Differences Between Wholes and Parts

Environmentalism often seem to motivate those who gave positive but equal values for water quality and biodiversity. But in many cases the respondent just did not differentiate between the whole and the part—this was about money to do good. This respondent was willing to pay $100 dollars for the chain and $100 for all lakes. "I believe in that. I could give $100 toward it...yes I felt good about that program and this was above and beyond my taxes. $100 is just sort of in my head as that would be, it, in other words the $100 had nothing to do with how much it would pay toward cleaning up a lake. It simply would be our family's maximum amount that we would ever put into a program outside of environmental things we're already involved in"

A person who was willing to pay $500 to protect biodiversity in Vilas and Onieda Counties and in all of Northern Wisconsin was not thinking about wholes and parts. "Well, I suppose I would associate it [biodiversity] with...just looking out in the backyard, you know, looking at the bird feeders and the deer feeders and ferns and the elms and the oaks making a place interesting, making it um I don't know, ecologically uh balanced is what I think about...In fact, this morning I saw a mother [deer] and two fawns out here...it was wonderful, and the hummingbirds were flying at the feeder at the same time, so here I'm looking at all these beautiful, wonderful things that make me feel just great, it's a spiritual connection to me."

Discussion

This paper began with the observation that the scope test is very appealing for its intuitive, common sense logic and theoretical simplicity. Our working conclusion after looking at aggregate and individual data across these four objects is that the scope test is neither a necessary nor sufficient criterion for judging the validity of values estimated using CV. Two of the objects showed scope, and two didn't. Those that showed scope, also showed cognitive and affective scope. But when you know more about the part, and like the part more you pay more for the part than the whole. Does this make the values invalid?

Are the WTP estimates for wolves invalid because people say they would pay no more for 800 than 300? If there were a simulated market for wolves, we are quite confident that those people who said they would pay zero, will actually pay zero. Moreover, we are certain that many people who said they would pay more for 300 wolves than 800 would actually do so with real money.

When we looked at the individual data we found the majority of the respondents failed to show scope for any of the four goods. This surely does not inspire confidence about the scope test as a validity criterion. Some people showed reverse scope for very good and well thought out reasons. So it is possible that some of the reverse scope judgments are validly representing what people would really be willing to pay if a market were available. So we believe that evidence of reverse scope is not sufficient to call the values in a single study invalid or the method in general flawed.

It does seem to us that very often the whole-part distinction is more in the mind of the analyst than in the mind of the respondent. For many of our respondents wolves went from an environmental good at 300 to an environmental bad at 800, and the Minocqua chain is a different attitude object than all of the lakes. While the chain is physically a subset of all the lakes, peoples ways of thinking about them may not be that simple. We think this might explain the some of the studies cited at the beginning of this paper failed to find scope. Environmental goods often have many attributes. Scope tests simply assume that one attribute (e. g. number of wolves, acres of wilderness etc.) defines the good, holding all other attributes constant. Respondents are smarter than that.

About biodiversity the public didn't have a clue. But they were willing to give very high values for what they thought biodiversity was. We fully expect that our respondents would have paid at least $173 dollars, on average, to keep the trees and plants and animals in their county. They wouldn't pay more for the whole north because the whole north is somewhere else. People will pay more for the part than the whole simply because they live in Vilas and Oneida counties or on the Minocqua Chain. Parts and wholes are two different things. More important what they said they were paying for was nature rather than what "biodiversity." The CV measure here even if it were validated by a simulated market, would be a most invalid indicator of the value of biodiversity as defined by the scientific community. Just because CV is sometimes badly applied, however, does not mean that the method itself is fundamentally flawed—just that it is difficult to use. The warning, "Don't try this at home," applies.

Even though spear fishing showed aggregate scope we were not able to predict the various scope types. Were other variables working or was there simply a lot of error variance here? Our guess is attitudes toward Indians and the impact of spearing on both the fish and the community would have explained these scope conditions but they were not asked on the surveys. Wolves did not show scope, but individual scope was quite predictable. We think the reason that we often got the same values for the whole and the part was that for most people 300 and 800 were really not very different. And those who did make the distinction between 300 and 800 felt that 800 was less desirable than 300. They didn't show scope but for good reason. We expect that the lack of distinction in numerical sense is why Desvouges et. al.(1993) did not find scope when respondents were presented with 2000, 20,000 and 200,000 birds. In each case the respondents were informed that the number presented was a relatively small proportion of birds.

There was certainly evidence in our data that the goods were loaded with ideological value as Kahnman observed in 1986. Willing to pay nothing for biodiversity or the same positive value for the whole and the part was tied to only one variable—environmentalism. When asked to give money for biodiversity the broad environmental value was the only predictor. But the fact that the values from the CV application are tied to this broad disposition doesn't mean that our respondents are unwilling to pay if we threatened to bulldoze forest and kill all the wild animals in Vilas and Onieda counties

So what to do? Based on these data we think that it is expensive and generally useless to have large split sample designs where half the people express a value for a part and half for a whole in an effort to show that a particular CV application is valid. Sometimes you will get scope, when people know and like the whole more than the part, but other times you won't, as when the whole and the part are really different goods, or people like the part more than the whole. The failure to find scope simply tells you something about the good, rather than the validity of the estimates, and the utility of the method. It would be nice if there were some easy statistical test like the scope test to compare groups to say conclusively if the CV application in any study was valid or not. Unfortunately, on the basis of our data, we don't believe the scope test *per se* is either necessary or sufficient to accomplish this task.

It would be better to focus the time and energy on trying to figure out a simulated market, even in a lab or an artificial setting for water quality, wolves, Indian spear fishing or biodiversity than to keep searching for evidence of validity or invalidity in scope tests.

Generally we think it is better to use CV on objects that individuals know a lot about, have strong feelings about, and lots of direct experience with. In these cases, whether one gets scope or not, the questions will make sense to the respondent and his or her answers will make sense. It is likely then that they will actually pay what they say they will pay. We think the values expressed about water quality and spear fishing have stronger cognitive and affective grounding, and in that sense would be more stable over time. The values for wolves although predictable and valid at the time we gathered our data could change if there were a key event associated with wolves. Our experience with biodiversity suggests that it is dangerous to think that elicited contingent values accurately represent people's values for abstract scientific constructs.

Reference List

Aaron, Henry J. 1994. "Distinguished Lecture on Economics and Government: Public Policy, Values, and Consciousness." *Journal of Economic Perspectives* 8(2):3-21.

Ajzen, Icek. 1987. "Attitudes, Traits, and Actions: Dispositional Prediction of Behavior in Personality and Social Psychology." *Experimental Social Psychology* 20:1-63.

Ajzen, Icek and George W. Peterson. 1988. "Contingent Value Measurement: The Price of Everything and the Value of Nothing?" Pp. 65-75 in *Amenity Resource Valuation: Integrating Economics With Other Disciplines*, (eds.) George L. Peterson, B.L. Driver, and Robin Gregory. State College PA.: Venture Publishing.

Allport, Gordon. 1935. "Attitudes." Pp. 789-844 in *Handbook of Social Psychology*, (ed.) C. Murchison. Worcester, MA: Clark University Press.

Bishop, Richard C., Patricia A. Champ, Thomas C. Brown, and Daniel W. McCollum. 1997. "Measuring Non-Use Values: Theory and Empirical Applications." Pp. 59-81 in *Determining the Value of Non-Marketed Goods*, Raymond J. Kopp, Werner W. Pommerehne, and Norbert Schwarz. Boston: Kluwer Academic Publishers.

Bishop, Richard C. and Thomas A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61:926-30.

Blamey, Russell K. 1998. "Decisiveness, Attitude Expression, and Symbolic Responses in Contingent Valuation." *Journal of Economic Behavior and Organization* 34:577-601.

Brown, Thomas C., Patricia A. Champ, Richard C. Bishop, and Daniel W. McCollum. 1994. "Which Response Format Reveals the Truth About Donations to a Public Good?" *Conference on Forestry and the Environment* (Banff, Alberta, Canada, Oct).

Cummings, R. G., D. S. Brookshire, and W. D. Schulze, (eds.). 1986. "Valuing Environmental Goods: An Assessment of the Contingent Valuation Method." Totowa, NJ: Rowman and Allenheld.

DeMaio, Theresa J. and Jennifer M. Rothgeb. 1996. "Cognitive Interviewing Techniques: In the Lab and in the Field." Pp. 177-95 in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, Norbert Schwarz and Seymour Sudman. San Francisco: Jossey-Bass Publishers.

Desvouges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson. 1993. "Measuring Natural Resource Damages With Contingent Valuation: Tests of Validity and Reliability." Pp. 91-159 in *Contingent Valuation: A Critical Assessment*, (ed.) Jerry A. Hausman. Amsterdam: North-Holland.

Diamond, Peter A. and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(4):45-64.

Dunlap, Riley E. and Kent D. Van Lier. 1984. "Commitment to the Dominant Social Paradigm and Concern for Environmental Quality." *Social Science Quarterly* 65:1013-28.

Dunlap, Riley and Kent D. Van Liere. 1978. "The "New" Environmental Paradigm." *Journal of Environmental Education* 9(4):10-19.

Eagly, Alice H. and Shelly Chaiken, (eds). 1993. *The Psychology of Attitudes*. New York: Harcourt Brace Jovanovich College Publishers.

Erevelles, Sunil. 1998. "The Role of Affect in Marketing." *Journal of Business Research* 42:199-215.

Fazio, Russell L. 1990. "Multiple Processes By Which Attitude Guide Behavior: The Mode Model As an Interative Framework." *Advances in Experimental Social Psychology* 23:75-109.

Fishbein, M. and I. Ajzen. 1975. *Belief, Attitude, Intention, and Behavior: an Introduction to Theory and Research.* Reading MA: Addison-Wesley.

Freeman, Myrick. 1993. *The Measurement of Environmental and Resource Values.* Washington D.C. Resources for the Future.

Frykblom, Peter. 1997. "Hypothetical Question Modes and Real Willingness to Pay." *Journal of Environmental Economics and Management* 34:275-87.

Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden. 1998. "Referendum Contigent Valuation, Anchoring and Willingness to Pay for Public Goods." *Resource and Energy Economics* 20:85-116.

Hausman, J. A., (ed.) . 1993. *Contingent Valuation: A Critical Assessment.* Amsterdam: Elsevier.

Heberlein, Thomas A. and Stanley J. Black. 1976. "Attitudinal Specificity and the Prediction of Behavior in a Field Setting." *Journal of Personality and Social Psychology* 33(4):474-79.

Irwin, Julie R. and Joan S. Spira. 1997. "Anomolies in the Values for Consumer Goods With Environmental Attributes." *Journal of Consumer Psychology* 6(4):339-63.

Kahneman, D. 1986. "Comments by Professor Daniel Kahneman." Pp. 185-94 in *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method*, (eds) R. G. Cummings, D. S. Brookshire, and W. D. Schulze. New Jersey: Rowman and Allanheld.

Kahneman, Daniel and Jack L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22:57-70.

Kim, Min-Sum and John E. Hunter. 1993. "Attitude-Behavior Relations: A Meta-Analysis of Attitudinal Relevance and Topic." *Journal of Communication* 43(1):101-42.

Kraus, Stephen J. 1995. "Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature." *Personality and Social Psychology Bulletin* 21(1):58-76.

Loomis, J. 1990. "Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques." *Journal of Environmental Economics and Management* 18:78-85.

Luzar, Jane E. and Kelli J. Cosse. 1998. "Willingness to Pay or Intention to Pay: The Attitude-Behavior Relationship in Contingent Valuation." *Journal of Social Issues* 27(3):427-45.

McGuire, William J. 1985. "Attitudes and Attitude Change." Pp. 233-346 in *The Handbook of Social Psychology*, 3rd ed. vol. 2, (eds.) Gardner Lindzey and Elliot Aronson. New York: Random House.

Milgrom, P. 1993. "Is Sympathy an Economic Value? Philosophy, Economics and the Contingent Valuation Method." *Contingent Valuation a Critical Assessment*, (ed.) J. A. Hausman. Amsterdam: Elsevier.

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

Mitchell, Robert C. and Richard T. Carson. 1995. "Current Issues in the Design, Administration, and Analysis of Contingent Valuation Surveys." Pp. 10-33 in *Current Issues in Environmental Economics*, eds. Per-Olov Johansson, Bengt Kristom, and Karl-Goran Maler. New York: Manchester University Press.

O'Connell, John F. 1982. *Welfare Economic Theory.* Boston: Auburn House Publishing.

Peterson, George L., B. L. Driver, and Robin Gregory, (eds.) . 1988. "Amenity Resource Valuation: Integrating Economics With Other Disciplines." Stage College, PA.: Venture Publishing Inc.

Petty, Richard E. and Jon A. Krosnick, (eds.). 1995. *Attitude Strength: Antecedents and Consequences*. Mahwah, N.J: Lawrence Erlbaum Associates.

Renata, Franc. 1999. "Attitude Strength and The Attitude-Behavior Domain: Magnitude and Independence of Moderating Effects of Different Strength Indices." *Journal of Social Behavior and Personality* 14(2):177-97.

Schkade, D. A. and J. W. Payne. 1993. "Where Do the Numbers Come From? How People Respond to Contingent Valuation Questions." *Contingent Valuation: A Critical Assessment* (Washington D.C. April 2-3), (ed.) Jerry A. Hausman. Washington D.C. Cambridge Economics Inc.

Schkade, David A. and John W. Payne. 1994. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation." *Journal of Environmental Economics and Management* 26:88-109.

Schuman, Howard and Michael Johnson. 1976. "Attitudes and Behavior." *Annual Review of Sociology*:161-207.

Schwarz, Norbert. 1997. "Cognition, Communication, and Survey Measurement: Some Implications for Contingent Valuation Surveys." Pp. 165-88 in *Determining the Value of Non-Marketed Goods*, eds. Raymond J. Kopp, Werner W. Pommerehene, and Norbert Schwarz. Boston: Kluwer Academic Press.

Spash, Clive L. and Nick Hanley. 1995. "Preferences, Information, and Biodiversity Preservation." *Ecological Economics* 12:191-208.

Stevens, Thomas H., Jaime Echeverria, Ronald J. Glass, and Tim Hager. 1991. "Measuring the Existence Value of Wildlife: What Do CVM Estimates Really Show?" *Land Economics* 67(4):390-400.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz, (eds.). 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco : Jossey-Bass Publishers.

Thiel, R. P. 1993. *The Timber Wolf in Wisconsin.* Madison: University of Wisconsin Press.

Willis, Gordan B., Theresa J. DeMaio, and Brian Harris-Kojetin. 1999. "Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques." Pp. 133-53 in *Cognition and Survey Research*, (eds.) Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau. New York: John Wiley and Sons.

Zajonc, R. B., Paula Pietromonaco, and John Bargh. 1982. "Independence and Interaction of Affect and Cognition." *Affect and Cognition: The Seventeenth Annual Carnegie Symposium on Cognition*, (eds.) Margaret Sydnor Clark and Susan Fiske. Hilsdale NJ: Lawrence Erlbaum Associates.
Zajonc, Robert B. 1980. "Feeling and Thinking: Preferences Need No Inference." *American Psychologist* 32(2):151-75.

Appendix A

Environmental CV Studies That Have Tested Scope Sensitivity*

Author(s)	Environmental Good(s) Studied	Scope Sensitivity?
Bowker & Didychuk (1994)	Agricultural Land in Canada	Yes
Boyle, Welsh and Bishop (1993)	Different CFS Water Flowages in the Grand Canyon	Yes
Carson, Mitchell & Ruud (1989)	Air Pollution—Visibility and Health Effects	Yes
Carson and Mitchell (1993)	National and Regional Freshwater Quality	Yes
Desvouges et. Al. (1992)	Waterfowl Deaths in Central Mississippi Flyway	No
Diamond et al. (1993)	Wilderness areas in four Rocky Mountain states	No
Duffield and Neher (1991)	Montana Waterfowl	Yes
Hoevenagel (1994)	Six environmental programs	Yes
Jakus (1992)	Gypsy Moth Control	Yes
Kahneman (1986)	Fish in Ontario Lakes	No
Kahneman and Knetsch (1992)	Environmental Services	Mixed
Krieger (1994)	Sport Fishing	Yes
Loomis et. Al. (1993)	Forests in Southeastern Australia	Mixed
Magnussen (1992)	North Sea Water Pollution Prevention	Yes
Mitchell and Carson (1986)	Drinking Water in the United States	Yes
Mitchell and Carson (1995)	The Kakadu Conservation Zone Australia	Yes
Mullarkey (1997)	Highway Expansion and Wetland Protection	Yes
Rowe et Al. (1991)	Northwest Oil Spills	Yes
Schkade and Payne (1994)	Waterfowl Deaths in Central Mississippi Flyway	No
Whitehead (1992)	North Carolina Sea Turtle Extinction	Yes
Whitehead & Blomquist (1991)	Kentucky Wetlands	Yes
Wu (1991)	Ohio Freshwater Streams	Yes

*Adapted from Carson(1997).

Valuing Benefits of Finnish Forest Biodiversity Conservation: Fixed and Random Parameter Logit Models for Pooled Contingent Valuation and Contingent Rating/Ranking Survey Data

Juha Siikamäki
Ph.D. Candidate
Department of Environmental Science and Policy
University of California, Davis
Email: juha@primal.ucdavis.edu

Presented at the Annual Meeting of Western Regional Project W-133
Miami, Florida
February 26-28, 2001

Acknowledgements: Doug Larson and Dave Layton are acknowledged for their comments and suggestions.

Abstract: This paper examines the use of contingent valuation and contingent rating/ranking valuation methods (CV and CR methods) in measuring willingness-to-pay (WTP) for nonmarket goods. Random parameter models are applied to CV and CR data and their performance is evaluated in comparison to conventionally used fixed parameter models. A framework for using data pooling techniques to test for invariance between separate sources of data is presented and applied to combined CV and CR data. The empirical application deals with measuring the WTP for conserving biodiversity hotspots in Finnish non-industrial private forests. Results suggest that the random coefficient models perform statistically well in comparison to the fixed parameter models that sometimes violate the assumptions of the conditional logit model. Based on the pooled models on combined data, parameter invariance between the CV and CR data cannot be uniformly accepted or rejected. Rejecting pooling of the data becomes more likely as more detailed response models are applied.

Introduction

This paper examines the use contingent valuation and contingent rating/ranking (CV and CR) methods in measuring willingness-to-pay (WTP) for nonmarket goods. Recent developments in discrete choice econometrics using random parameter models are applied to CV and CR data, and their performance is evaluated in comparison to conventionally used fixed parameter econometric models. Further, invariance between the CV and CR data is examined by data pooling techniques.

Stated preference methods (SP methods) are widely used in measuring economic values related to the environment. Standard SP applications include conducting surveys, in which respondents are described hypothetical alternatives, usually policy options. Each policy option results a certain supply of nonmarket good, such as environmental quality, for certain costs to respondents. The respondents are asked to evaluate the alternatives and state their preferences regarding them. The CV is based on asking for acceptance/refusal of hypothetical payment for implementing a policy alternative; the CR relies on asking respondents to rate or rank the available alternatives, at the simplest by choosing a preferred alternative. Obtaining responses for a variety of cost-environmental quality combinations, data with implicit information on individual tradeoffs between money and environmental quality are collected. The tradeoffs can be quantified by using discrete choice econometric models, that explain the observed choices by attributes of policy alternatives and respondents. In essence, the econometric models are used to measure an individual level exchange rate between a nonmarket good and money. Willingness to pay (WTP) for changes in the environmental quality can then be calculated using the estimation results.

Although several stated preference methods are currently in use, their performance and consistency has not been exhaustively studied. Examples of studies on differences between SP methods include Desvouges and Smith 1983, Magat et al. 1988, Boxall et al. 1996, and Stevens et al. 2000. They all suggest substantial differences between the various SP methods. However, the objective of all the methods is to measure essentially the same tradeoffs between money and changes in the environmental quality, and their results should therefore be very similar.

Previous studies on differences across the SP methods are typically based on fixed parameter discrete choice models, usually logit models. The assumptions and properties of fixed logit models are restrictive, but more flexible models with random parameters have been practically unavailable due to limitations in computing power and simulation based econometric techniques. Both constraints have recently been greatly relaxed and random parameter models are now possible to be employed in modeling the discrete choice SP data. Both Train's (1998) analysis of recreational fishing site choice and Layton's (2000) work on rankings data demonstrate that random parameter formulation can substantially improve the statistical performance of the econometric models typically used for valuing natural resources.

As mentioned, the differences between the SP methods have been previously studied by relatively restrictive models. Hence, it is justified to re-examine the differences of the SP methods by using less restrictive random parameter models. More flexible models let us evaluate if the previous conclusions have resulted from actual inequalities between different SP data sources, or perhaps from using overly restrictive econometric models.

Adamoviz et al. (1994) tested for differences between the observed and stated choices by estimating models for combined data on observed and stated choices. Recently, Hensher et al. (1999) provide a general framework for applying data pooling techniques to test for the invariance between separate sources of data. The data pooling approach is adopted here and used in testing for the equality of the CV and CR data. The approach can be easily extended into different settings.

The data pooling approach enables comparing different data sources already in estimation. Several benefits follow in comparison to the traditional approach of estimating separate models for different data sets and comparing their results afterwards: First, likelihood ratio based tests for data source invariance become available. Second, if data from different SP sources can in fact be considered equal, the pooled econometric models provide practical means to utilize all the information in data collected. This in turn can result in more reliable model estimates than the unpooled models.

The empirical application deals with measuring WTP for conserving especially valuable habitats (biodiversity hotspots) in Finnish non-industrial private forests. According to ecologists, protection of the biodiversity hotspots is particularly important for biodiversity conservation in Finland. The hotspots cover a total of 1.1 million hectares, that is some 6 % of the Finnish forests. Current regulations protect some 110,000 hotspot hectares and extending their protection is currently debated. This study evaluates the potential conservation policy alternatives by examining public preferences for them.

Forest conservation in Finland is an inexhaustible source of public debates and policy conflicts. Clearly, management and harvesting of forests are the primary reasons for species extinction. Rather intensive forest management practices over a long period of time have provided country with more timber resources than ever in the known past. At the same time, substantial losses of old forests and other important habitats for many currently threatened species have resulted. On the other hand, a big share of the country's exports consist of forest products such as paper- and sawmill products. Economic interests related to forests are therefore evident. Noting further that forests consist mostly (65-75 %) of small holdings (avg. size 100 acres), owned by private households, and that almost 10 % of the Finnish population owns some areas forests, it is clear that forest conservation policies are both of considerable public and regulatory interest.

The specific objectives of this paper are to

1) Review and to discuss current logit models for the CV and CR data.
2) Examine the random parameter modeling approach in comparison to fixed parameter models.
3) Test for differences between SP methods by using data pooling methods.
4) Analyze WTP estimates for both fixed and random parameter, and unpooled and pooled models for CV and CR data.

The rest of the paper is organized as follows: The first section represents the econometric models for CV and CR data, including fixed and random coefficient models. The next section explains how data pooling techniques can be used to test for data invariance between different stated preference data sources. The empirical section starts with a description of the public survey for preferences for biodiversity conservation in Finland. Results start with fixed logit models and continue with results for random parameter models. After separate estimation of the CV and CR data, the two data sets are pooled and invariance between the CV and CR data tested. The results section is concluded with the WTP estimates for different models. Last, the results of the study are discussed and concluded.

Econometric Models for Contingent Valuation and Contingent Rating/Ranking Survey Responses

Econometric models for stated preference surveys are typically based on McFadden's (1974) random utility model (RUM). The following section uses the RUM as a point of departure for explaining various econometric models for CV and CR survey responses. The CV section draws from works by Hanemann (1984), Hanemann et al. (1991), and Hanemann and Kanninen (1996); the CR section relies on McFadden (1974), Beggs et al. (1981), Chapman and Staelin (1982), Hausman and Ruud (1987), and on recent works by Train (e.g. 1998), Train and McFadden (2000) and Layton (2000a).

Random Utility Theoretic Framework for Modeling Individual Choices

Typical stated preference surveys measure individual tradeoffs between changes in environmental quality q and costs A of implementing them. This is accomplished by asking respondents to state their choices between the status quo with zero cost and one or more hypothetical policy alternatives with altered environmental quality and its costs. Consider an individual i choosing a preferred alternative from a set of m alternatives, each alternative j providing utility U_{ij}, that can be additively separated into an unobserved stochastic component ε_{ij} and a deterministic component $V_{ij}(q_j,y-A_j)$ i.e the restricted indirect utility function that depends only on individual's income y and environmental quality q. The utility of alternative j can then be represented as

$$U_{ij}=V_{ij}(q_j,y-A_j)+\varepsilon_{ij} \tag{2.1}$$

The stochastic ε_{ij} represents the unobserved factors affecting the observed choices. They can be related to individual tastes, choice task complicity, or any other factors with significant influence on choices. They are taken into consideration by individual j choosing between the alternatives, but to an outside observer, ε_{ij} remains unobserved and stochastic in the econometric modeling. From the viewpoint of individual making a choice, utility has no stochastic nature.

Choices are based on utility comparisons between the available alternatives, and the alternative providing the highest utility becomes the preferred choice. The probability of person i choosing alternative j among all the m the alternatives therefore equals the probability that the alternative j provides person i with greater utility U_{ij} than any other available alternative with U_{ik}. It is determined as

$$P_{ij} = P\,(U_{ij} > U_{ik},\, k = 1, \ldots, m, \quad k \quad j), \tag{2.2}$$

Denoting the difference of random components between alternatives j and k as $\varepsilon_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$, and the difference between the deterministic components as $\Delta V_{ijk}\,(.)= V_{ik}\,(q,y-A_k)- V_{ij}(q,y-A_j)$, the probability P_{ij} can be presented as probability

$$P_{ij} = P\,(\varepsilon_{ijk} > \Delta V_{ijk}(.),\, k = 1, .., m, \quad k \quad j) \tag{2.3}$$

Estimating parametric choice models requires specification of both the distribution of ε_i and the functional form of V_{ij}. Specification of ε_{ijk} determines the probability formulas for the observed responses; the functional form of V_{ij} is employed in estimating the unknown parameters of interest. Denoting all the exogenous variables of alternative j for the ith person as a vector X_{ij}, and the unknown parameters as a vector β, V_{ij} is typically specified as linear in parameters $V_{ij}=X_{ij}\beta$.

The following sections describe response probability formulas for different contingent valuation (CV) and contingent ranking (CR) models. Response probability formulas can be thought of as a likelihood function for the ith person. Since observations are independent, the likelihood function for the total sample is simply a sum of individual likelihood functions. The total maximum likelihood function can then be employed in estimating the V_{ij}.

Logit models for CR data

Assume in the following that random terms ε_j and ε_k are independently and identically distributed, type I generalized extreme value (GEV) random variables. It follows that their difference ε_{ijk} is logistically distributed. Under these assumptions, McFadden (1974) showed that choice probability P_{ij} in (2.5) is determined as a conditional logit model

$$P_{ij} = \frac{e^{\mu X_{ij}\beta}}{\sum_{k=1}^{k=m} e^{\mu X_{ik}\beta}} \tag{2.4}$$

The log-likelihood function for conditional logit model is

$$\ln L = \sum_{i}^{N} \ln P_{ij} \tag{2.5}$$

The parameter μ is a scale factor that appears in all the choice models based on RUM. It links the structure of random terms and the parameter estimates of $V_{ij}=X_{ij}\beta$. With data from a single source, the scale factor is typically set equal to one, left out, and parameter vector β estimated given the restricted scale factor. This is necessary for identification; without the imposed restriction on μ, neither μ nor β could be identified. However, in combining data from different sources, the scale factor plays an essential role. Since pooling of CV and CR data plays an important role in the analysis, scale parameters are included in all the following models. The role of the scale factor in pooling different sources of data will be discussed more in section 2.5.

Beggs et al (1981) and Chapman and Staelin (1982) extended the conditional logit model to modeling ranking of alternatives. A rank-ordered logit model treats ranking as *m-1* consecutive conditional choice problems. It assumes that ranking results from *m-1* utility comparisons, where the highest ranking is given to the best alternative (the preferred choice from the available alternatives), the second highest ranking to the best alternative from the remaining *m-1* alternatives, third from the remaining *m-2* alternatives, and so on. The probability of the observed ranking *r* for the person *i* is given by

$$P_{ir} = \prod_{j=1}^{m-1} \frac{e^{\mu X_{ij}\beta}}{\sum_{k=j}^{k=m} e^{\mu X_{ik}\beta}} \tag{2.6}$$

Hausman and Ruud (1987) developed a rank-ordered heteroscedastic logit model that is flexible enough to take into account possible increases (or decreases) in variance of the random term in the RUM as the ranking task continues. It is based on formulation with a rank-specific scale parameter that accounts systematic changes in the variance of the random term. By its structure, a rank-ordered heteroscedastic logit model can identify *m-2* scale parameters.

$$P_{ir} = \prod_{j=1}^{m-1} \frac{e^{\mu_j X_{ij} \beta}}{\sum_{k=j}^{k=m} e^{\mu_j X_{ik} \beta}} \tag{2.7}$$

As with the conditional logit model, the log-likelihood function for rank-ordered logit models (2.6) and (2.7) is the sum over individual probabilities over the whole sample.

Logit models for CV data

A single bounded discrete choice CV method is based on asking respondents if they would or would not be willing to pay certain reference amount *Bid* of money for altering the environmental quality q. Data consist of binary responses that result from yes/no answers to CV questions, asking for refusal/acceptance of paying an amount *Bid* for some policy alternative. In essence, the CV-method asks respondents to choose between status quo with utility $U_{i0}(q_0) = V_{i0}(q_0) + e_{i0}$ and an alternative providing utility $U_{i1}(q_1) = V_{i1}(q_1, y\text{-}Bid) + e_{i1}$. Given a logistically distributed stochastic term in the RUM, the probability of individual i choosing the alternative with costs *Bid* and environmental quality q_1 is the probability of obtaining a *Yes*-answer from person i. Expressing the observed parts of utilities as $V_{i0} = X_{i0}$ and $V_{i1} = X_{i1}$, the probability of a Yes-answer is given by the conditional logit model with two alternatives.

In double bounded CV, respondents are asked a follow-up question based on the first response. The objective is to gather more information on WTP than is possible by asking just a single question. Respondents who answered *Yes* to the first question (*FirstBid)* are asked a similar second question, this time with *HighBid* > *FirstBid.* Respondents who answered *No* get a second question with *LowBid* < *FirstBid.* Second responses provide more detailed data on individual preferences between the two alternatives and the choice probabilities can now be determined based on responses to two separate questions. Four possible response sequences can be observed: *Yes-Yes*, *Yes-No*, *No-Yes* and *No-No.* Using the conditional logit model, and denoting the exogenous variables for questions with *FirstBid, HighBid* and *LowBid* by X_{iFB}, X_{iHB} and X_{iLB}, the probabilities of the different responses are given by:

$$\begin{aligned}
&\text{P(Yes-Yes)} = P_i\,(YY) = \frac{e^{\mu X_{iHB}\beta}}{e^{\mu X_{iHB}\beta} + 1} \\
&\text{P(Yes-No)} = P_i\,(NY) = \frac{1}{1+e^{\mu X_{iHB}\beta}} - \frac{1}{1+e^{\mu X_{iFB}\beta}} \\
&\text{P(No-Yes)} = P_i\,(NY) = \frac{1}{1+e^{\mu X_{iFB}\beta}} - \frac{1}{1+e^{\mu X_{iLB}\beta}} \\
&\text{P(No-No)} = P_i\,(NN) = \frac{1}{1+e^{\mu X_{iLB}\beta}}
\end{aligned} \tag{2.9}$$

Using dummy variables I_{yy}, I_{yn}, I_{ny}, I_{nn} to indicate *Yes-Yes*, *Yes-No*, *No-Yes* and *No-No* responses, the log-likelihood function for double-bounded CV is

$$L = \sum_{i=1}^{n} \ln[I_{yy}P_i(YY) + I_{yn}P_i(YN) + I_{ny}P_i(NY) + I_{nn}P_i(NN)] \qquad (2.10)$$

Random parameter logit models

Although typically applied to SP data, some undesirable properties and assumptions are embodied in the fixed parameter logit models. First, they overestimate the joint probability of choosing close substitutes. This is known as the Independence of Irrelevant Alternatives (IIA) property (McFadden 1974). Second, they are based on the assumption that the random terms ε_{ij} are independently and identically distributed, although in practice it is likely that individual specific factors influence evaluation of all the available alternatives and make random terms correlated instead of independent. Third, assuming homogeneous preferences alone is restrictive. Any substantial variation in individual tastes conflicts with this assumption, possibly resulting in violations in many applications.

Random parameter logit (RPL) models have been proposed to overcome possible problems of the fixed parameter choice models (e.g. Revelt and Train 1998, Train 1998, Layton 2000). The RPL is specified similarly as the fixed parameter models, except that the parameters β now vary in the population rather than stay the same for everybody. Utility is expressed as a sum of population mean b, individual deviation , which accounts for differences in individual taste from the population mean, and an unobserved i.i.d. random term ε. Total utility for person i from choosing the alternative j is determined as

$$U_{ij} = X_{ij}b+X_{ij}\ \ _i+\varepsilon_{ij} \qquad (2.11)$$

where $X_{ij}b$ and $X_{ij}\ \ _i+\varepsilon_{ij}$ are the observed and unobserved parts of utility. Utility can also be expressed in form $X_{ij}(b+\ \ _i)+\varepsilon_{ij}$, which is easily comparable to fixed parameter models. The only difference is that previously fixed β now varies across people as $\beta_i=b+\ \ _i$.

Although the RPL models account for heterogeneous preferences via parameter $_i$, individual tastes deviations $_i$ are neither observed nor estimated. The RPL models aim at finding the different moments, for instance the mean and the deviation, of the distribution of β, from which each β_i is drawn. Parameters β vary in population with density $f(\beta|\ \ )$, with denoting the parameters of density. Since actual tastes are not observed, the probability of observing a certain choice is determined as an integral of the appropriate probability formula over all the possible values of β weighted by its density. Probability for choosing alternative j out of m alternatives can now be written as

$$P_{ij} = \int \left[\frac{e^{\mu X_{ij}\beta_i}}{\sum_{k=j}^{k=m} e^{\mu X_{ik}\beta_i}} \right] f(\beta \mid \Omega) d\beta \qquad (2.12)$$

Equation (2.12) is the random parameter extension of the conditional logit model (2.4). Random parameter models for the rank-ordered logit models and the double bounded CV are defined similarly. Extension is straightforward and not replicated here. It suffices to note that they are formulated by replacing the bracketed part of (2.12) by the appropriate probability formula.

Integral (2.12) cannot be analytically calculated and must be simulated for estimation purposes. Therefore, exact maximum likelihood estimation is not available and simulated maximum likelihood is to be used instead. Train has developed a method that is suitable for simulating (2.12) and its many extensions needed in this study. His simulator is smooth, strictly positive and unbiased (Brownstone and Train 1999), and can be easily modified to allow for non-negative/positive random parameters. That is particularly practical in CV and CR studies in which theoretical considerations often suggest restrictions for parameter values. Simulating (2.12) is carried out simply by drawing a random β_i, calculating the bracketed part of the equation, and repeating the procedure over and over again. Although Train's simulator is unbiased for just one draw of β_i, its accuracy is increased with the number of draws. Using R draws of β_i from $f(\beta|\ )$, the simulated probability of (2.12) is

$$SP_{ij} = \frac{1}{R}\sum_{r=1}^{r=R}\frac{e^{\mu X_{ij}\beta_{ir}}}{\sum_{k=j}^{k=m} e^{\mu X_{ik}\beta_{ir}}} \qquad (2.13)$$

Simulator (2.13) can be extended to rank-ordered logit model and to logit models for single and double bounded CV. The only required change is replacing the R times summed portion of (2.13) with the rank-ordered or double bounded CV probability formulas, as expressed by (2.6), and (2.9). In estimating mean and variance for the distribution of β, (2.17) can be employed by defining $X_{ij}\beta_{\ ir}=X_{ij}(b+\ e_{ir})$, where b and are estimated mean and deviation parameters and e_{ir} a standard normal deviate for rth replication for individual i. Estimation of parameters is carried out by maximizing the simulated likelihood function, determined by the appropriate simulated response probability formula, in much the same way as for fixed logit models. The simulated log-likelihood function for the random parameter conditional logit model (2.13) is

$$SL = \sum_{i}^{N} \ln SP_{ij} \qquad (2.18)$$

It is worth noting that RPL models are also flexible in approximating the response probabilities generated by other than Type 1 GEV distributed random terms in the RUM, such as normally distributed random terms. McFadden and Train (2000) show that any discrete choice model derived from random utility maximization can be approximated arbitrarily close by random parameter multinomial logit model.

Pooling Data

The scale of estimated parameters in all the choice models based on the RUM is related to the magnitude of the random component in the RU model. The scale factor μ relates the estimates with the random component, being inversely related to the variance of the random component in the RUM. Using a single source of data, μ is typically set equal to one since it cannot be identified. The estimated vector of coefficients β is therefore confounded with constant μ. This in turn makes absolute values of the parameter estimates incomparable between different data sets; only the ratios of coefficients are comparable across different sources of data (Swait and Louviere 1993).

Consider n separate sources of stated preference data, such as survey data using CV and CR. Normalizing scale factors equal to one in estimation of separate data sources, each data $q=1,...,n$ provides us with parameter estimates β_q. Denoting the scale parameters of different data sources with μ_q, n vectors $\mu_q\beta_q$ of parameter estimates results. Pooling n sources of data, it is possible to identify $n-1$ scale parameters for different data sources. Fixing one scale factor, say $\mu_1=1$, the rest $n-1$ estimated scale parameters are inverse variance ratios relative to the reference data source (Hensher et al. 1999).

Denote the vector of CV and CR estimates by $\mu_{CV}\beta_{CV}$ and $\mu_{CR}\beta_{CR}$. Pooling the CV and CR models, fixing $\mu_{CV}=1$, and estimating μ_{CR}, then accounts for possible differences in the variance of random terms between the CV and CR data. To test for the parameter invariance between the CV and CR data, models with and without restriction $\beta_{CV}=\beta_{CR}$ need to be estimated. Likelihood ratio tests can then be applied to accept/reject the imposed parameter restriction. If the null hypothesis cannot be rejected, the data generation processes can be considered generated by the same taste parameters but still have variance differences. Restricting both $\beta_{CV}=\beta_{CR}$ and $\mu_{CR}=1$ provides an even stricter test of data invariance, testing for both parameter and random component invariance. If not rejected, the two data sets can be considered similar and absolute parameter estimates comparable across the source of data.

Data

Data were collected using a mail survey, sent out in spring 1999 to a sample of 1740 Finns between 18-75 years of age. The sample was randomly drawn from the official census register, and divided into two random sub-samples of 840 and 900 respondents. The first sub-sample received a double bounded CV questionnaire and the second sub-sample a CR questionnaire.

Questionnaires started with questions about respondents attitudes on how important the different aspects of forests, such as their economic importance and different uses (timber production, recreation, nature conservation etc) should be in formulating forest policy. Next, respondents were asked to state how important issues such as public healthcare, education, employment, economic growth, nature conservation and equal income distribution should be in formulating public policies in general. Thereafter, respondents were asked still a number of attitude questions about forest conservation, landowners' and public's responsibilities in conservation, and the acceptability of forcing landowners to protect forests by regulatory approaches. The next section of the questionnaire included the valuation questions, described in more detail later. The questionnaire concluded with questions on the respondent's socioeconomic background.

While designing the survey, questionnaire versions went through several rounds of modifications and reviews by accustomed SP-practitioners, as well as other economists, foresters and ecologists with expertise in survey methods and/or biodiversity conservation. After hearing their comments, questionnaires were tested by personal interviews and a pilot survey (n=100), and modified based on the results. The final survey was mailed out in May 1999. A week after the first mailing, everyone in the sample was sent a reminder card. Two more rounds of reminders with a complete questionnaire were sent to non-respondents in June-July. The CV and CR surveys resulted in 48.9 % and 50 % response rates, respectively. After censoring for all the missing answers to valuation questions, 376 CV and 391 CR responses were available for further examination.

WTP is measured for three hypothetical conservation programs: Increasing conservation from the current 120,000 hectares to (1) 275,000 hectares, (2) 550,000 hectares and (3) 825,000 hectares. The new alternatives correspond to protection of 25 %, 50 % and 75 % of all the available biodiversity hotspots. In designing the survey, special attention was paid to formulating conservation policy scenarios so that they were policy relevant, credible and easy to understand. A one page easy to read section in the questionnaire explained different conservation programs and their details.

The CR survey described to respondents the status quo and all three hypothetical programs of setting aside additional 155,000, 430,000 and 705,000 hectares of hotspots for 30-years[1]. Table 1 describes how the conservation programs were summarized in the CR questionnaire. Using a scale from 0 to 10, each respondent was asked to rate the four programs. Note here that the respondents are not asked to hypothetically buy any forest areas; they are simply asked to express their preferences regarding different conservation programs that would each result in different conservation levels and costs to their households. The three hypothetical programs were assigned costs using the same variation across the respondents and the conservation programs as in the CV survey, described in more detail later.

Table 1. Four possible conservation projects presented for the CR respondents

Conservation Project	Total area under conservation	Proportion of conserved of all the Finnish forests
1. Current regulation, no new conservation	120,000 hectares	0.6 percent
2. Increasing conservation to cover one fourth (25%) of the biodiversity hotspots	275,000 hectares	1.5 percent
3. Increasing conservation to cover half (50%) of the biodiversity hotspots	550,000 hectares	3 percent
4. Increasing conservation to cover three fourths (75%) of the biodiversity hotspots	875,000 hectares	4.5 percent

The respondents of the CV questionnaires were divided into two groups. The first group was asked to state their WTP for the first two policy alternatives, i.e. 275,000 and 550,000 hectares as described in Table 1, and the second group for the 550,000 and 825,000 hectare alternatives. Each respondent was asked two separate CV questions, and responses for 50 % conservation were therefore collected by both the first and the second WTP questions, depending on the respondent's sub-sample. The CV method was applied using a double bounded format. The bid vector in the CV survey consisted of first bids between US$ 4-500, and the follow-up bids between US$ 2-800, with seven different starting bids. The same bid amounts appeared a first and second bids for different respondents, and the bids for different levels of conservation were randomly chosen from across the full vector of bids.

The final survey consisted of 29 different questionnaire versions; 14 were CV surveys and 15 CR surveys. In both types of surveys, WTP was measured as an increase in the annual tax

[1] The length of protection is determined as 30 years because the current policy programs for voluntary conservation are based on 30 year protection.

burden of the household. Except for the valuation question, the CV and CR questionnaires were similar. The set up for the data collection is such that only the choice task in the valuation question vary between the CV and CR respondents.

Results

The next sections report and discuss the results of fixed and random parameter logit models separately for the CV and CR data. After separate estimation of the CV and CR models, the two data sets are combined and invariance between the CV and CR data tested. All results are based on maximum likelihood estimation of the models described earlier in this paper. They were programmed and estimated in GAUSS.

Estimated models use a dummy specification for the conservation programs. In other words, conservation was modeled as three different conservation programs, not as a continuous variable of conserved hectares under each policy alternative. This results in several benefits: First, the specification is very flexible and does not restrict the value function for conservation to follow any certain functional form, letting it take practically any form instead. Second, the WTP for different conservation programs can now be calculated simply as ratios of the estimated parameters. Third, separate dummies can be used in measuring the variances of taste parameters for different extents of conservation. The observed part of RUM is estimated as

$$V_{ij} = \beta_{BID}BID_{ij} + \beta_{D25}D25_{ij} + \beta_{D50}D50_{ij} + \beta_{D75}D75_{ij} \tag{2.19}$$

where BID is the annual cost to the respondent's household from implementing policy alternative j, and D25, D50 and D75 dummy variables that indicate the extent (25 %, 50 % and 75 %, respectively) of conservation in policy alternative j. The specification of the V_{ij} stays the same throughout the reported models.

Fixed Parameter Logit Models

Contingent Valuation

The CV data with 376 observations was first censored for missing responses. The remaining 306 observations with complete double bounded responses to the both WTP questions were employed in the estimation.

As mentioned, the CV sample was divided into two groups, with 50 % conservation program in either the first or the second WTP question. Differences between the responses from these two groups were studied by first estimating separate parameters for 50 % conservation for the two groups. The parameters were then restricted equal, and a constrained model with a single parameter for the 50 % alternative was estimated. Based on the unconstrained and constrained model results, a likelihood ratio test was formulated to test for the similarity of responses to 50 % conservation program between the two groups.

Table 2. Model estimates for contingent valuation data

Model:	Fixed logit CV	
Estimate (|*t*-statistic|)	Unconstrained	Constrained (*D50_1=D50_2*)
BID	-0.3527	-0.3529
	(13.055)	(13.065)
D25	1.5996	1.6002
	(6.509)	(6.511)
D50_1	1.3864	
	(5.287)	
D50_2	1.3066	
	(5.225)	
D50		1.3445
		(7.048)
D75	0.9640	0.9645
	(3.610)	(3.612)
Mean LL	-380.771	-380.799
LL at 0 [a]	-446.775	-446.775
Pseudo R^2	0.148	0.148

[a] Pseudo R^2 is calculated as 1-LLU/LLR, where LLU and LLR are the log-likelihood values for the estimated model and model with only a constant.

Note: Number of observations 306. Table cells for structurally non-identified parameters are shaded in this and the following tables.

Table 2 reports the results of the fixed logit models for CV data. Dependent variable in the models is the probability of Yes-answer to the dichotomous choice WTP question. The estimated parameters are defined as follows: *BID*[2] is the household's annual cost from implementation of the suggested conservation program; *D25* and *D75* are dummies that indicate the 25 % and 75 % levels of conservation in the WTP question; *D50_1* stands for the 50 % conservation program as the first WTP question; *D50_2* stands for the 50 % conservation program as the second WTP question; *D50* is a dummy that pools 50 % conservation programs by restricting *D50_1 = D50_2.*

Both double bounded models result in highly significance parameter estimates. Estimates for the *D25* are significantly greater than zero, suggesting that the 25 % conservation program is preferred to status quo. The estimates for *D50* and *D75* are positive and greater than zero. Therefore, they are also preferred to status quo. However, the estimate for *D50* is systematically lower than the estimate for *D25*, and the estimate of *D75* in turn lower than the one for *D50.* This suggests a conservation policy preference order (25 % > 50 % > 75 % > status quo), thereby an increasing WTP from status quo to 25 % conservation, an possibly a negative marginal WTP for the higher levels of conservation.

No statistically significant differences between the responses to the 50 % conservation program are found based on the values of the maximized log-likelihood functions. The likelihood

[2] Note that in all the results reported in this chapter, variable BID is divided by 100 to facilitate estimation.

ratio testing[3] for the constraint *D50_1=D50_2* results in LR test static value 0.05. The test therefore rejects the null hypothesis of different parameter estimates for the *D50_1* and *D50_2*, and suggest accepting the constrained model. Based on strong rejection of the null hypotheses, the constrained models are used in the further analysis. In practical terms, this means pooling the responses to the 50 % conservation program and estimating *D50* as a single parameter.

Contingent Rating/Ranking

All the models for the CR data are based on rankings that were obtained by transforming respondents' ratings for policy alternatives into a preference ordering, assuming that preferred alternatives were rated higher than the less preferred ones. Rankings utilize only ordered information on preferences. Respondents with ratings sequences (3,2,1,0) and (10,9,3,1) are therefore considered similar responses with the same preference ordering A>B>C>D. In building the ranking data, observations with ties or missing ratings were censored, leaving a total of 270 observation left for the estimation. The results are therefore based on data with full and unique rankings of all four policy alternatives.

The specification of the CR models is the same as for the CV data. The following models were estimated: (1) a conditional logit model for the highest ranked alternative out of all the alternatives i.e. conditional logit model for preferred choice, (2) rank-ordered logit models for 2 and 3 ranks, both as rank-homoscedastic (ROL) and rank-ordered heteroscedastic models (ROHL). The models for 2 ranks explain the first two preferred alternatives; the model for 3 ranks a full ranking of the four alternatives.

Several rank-ordered logit models were estimated in order to examine the consistency of rankings. Information on more than only the preferred alternative is valuable, but beneficial only if the rankings are consistent and generated by the same parameters (e.g. Layton 2000). It is known that the variance of stochastic term in RU model tends to change (typically increase) as the ranking continues. This has been suggested to result from the respondents ranking the preferred alternatives with more care than the less preferred alternatives, causing data on 2nd ranks to be more noisy than data on 1st rank, data on 3rd rank to be more noisy than data on 2nd rank, and so forth.

The changing variance of random term between the ranks violates the i.i.d. assumption of rank-ordered logit model. Inconsistent rankings reveal violations of the assumption. Models with violations should be rejected and models for fewer ranks used instead. If variance of the random term changes sufficiently systematically and similarly over the rankings, the problems caused by the inconsistency of rankings could be solved by employing a Hausman–Ruud rank-ordered heteroscedastic logit model. Testing for the consistency of for instance the first and the second ranks can be carried out by estimating separate models for the first and the second ranks. These models result in two maximized log-likelihood values, denoted by LL_1 and LL_2. Constrained model is then estimated as a rank-ordered logit model for two ranks, resulting in a maximized log-likelihood value LL_R. A LR-test statistic with degrees of freedom equal to the number of constrained parameters is calculated as $-2*(LL_R-LL_1+LL_2)$. If the test statistic is insignificant, the ranks can be pooled and a rank-ordered logit model for two ranks used. If the LR-test fails to accept the pooling of the ranks, a Hausman-Ruud style rank-heteroscedastic logit model can be estimated and similar LR-test procedure carried out using it; this time with one less degrees of freedom because an additional parameter is estimated.

[3] LR-test statistic is calculated as –2(LLR-LLU), where LLR and LLU are the values of maximized log- likelihood function for constrained and unconstrained models, respectively (e.g. Amemiya 1983).

Table 3. Fixed logit models for contingent rating/ranking data

Model:	1 Rank	2 Ranks		3 Ranks	
Estimate (|*t*-statistic|)		ROL	ROHL	ROL	ROHL[a]
BID	-0.0888	-0.0779	-0.0484	-0.0592	-0.00099
	(-3.153)	(-3.904)	(-2.275)	(-3.635)	(-1.340)
D25	0.3537	0.7561	0.6217	1.1588	0.4519
	(1.328)	(3.783)	(3.321)	(6.282)	(2.083)
D50	0.2760	0.5559	0.4358	0.9959	0.3445
	(0.818)	(2.242)	(2.226)	(4.522)	(2.000)
D75	0.2055	0.1155	0.0672	0.1473	0.0480
	(0.507)	(0.368)	(0.297)	(0.531)	(0.578)
$\mu_{Rank2,3}$			1.8146		4.6160
			(2.765)		(1.980)
LL	-150.709	-261.938	-260.515	-319.496	-307.919
LL at zero	-162.556	-291.379	-291.379	-372.657	-372.657
Pseudo R^2	0.073	0.101	0.106	0.143	0.174

Note: Number of observations 270.
[a] Hausman-Ruud rank-heteroskedastic model for three ranks is estimated with common scale factor for the second and third rank. Estimating separate scale factors for second and third ranks is structurally possible but they could not be identified.

Table 3 reports the results of fixed parameter CR models. The signs, relative magnitudes and statistical significance of parameter estimates are rather well in line with the estimates for the CV models. The parameter estimates for *D25* are always greater than the estimates for *D50* and *D75*. The CR models also suggest uniformly that in the average, the 25 % conservation policy is preferred over the other policy alternatives, including status quo. Moreover, the relation between the *D50* and *D75* is similar as in the CV results, with the *D75* estimates having the smallest absolute but still positive estimates.

The insignificance of *D75* estimates in all the CR models is distinctive in comparison with the CV results. This could be related to the questionnaire design; the 75 % conservation alternative was always presented as the last policy alternative, possibly resulting in less careful rating than for the first, second and third policy alternatives. Another possibility is that preferences regarding the 75 % conservation policy are simply so heterogeneous that the identification of parameters is troublesome. Conditional logit model for the first rank results in statistically insignificant estimates for all the parameters except for the *BID*. In addition to the noise in the data, this can be related to relatively small number of observations.

Examining next the homoscedastic rank ordered logit models for 2 and 3 ranks, it is noted that all the statistically significant parameter estimates are greater in absolute magnitude than their counterparts in the first rank model. This is logical; by utilizing more information on the individual preferences, relative magnitude of the stochastic term in the RUM is decreased and substituted by higher parameter estimates and therefore higher proportion of observed variation. Further, all the parameter estimates except *D75* are now statistically significant both in the 2 and 3 rank models. The pseudo R^2 measures are still relatively low, although higher than for the first rank model.

Although exploiting information on more than only the first rank first seems to provide improvements compared to first rank model, the consistency of rankings is necessary to be examined before accepting the rank ordered models. The LR-test results for the consistency of rankings are reported in Table 4. The tests suggest that both the homoscedastic logit models (ROL) for 2 ranks can be accepted and the first two ranks pooled. Evidence is not particularly strong but the LR-test statistics are insignificant at the 1 % level. Consistency of three ranks is rejected with strong statistical evidence.

Table 4. Hypothesis tests on pooling different ranks

	2 Ranks		3 Ranks	
	ROL	ROHL	ROL	ROHL
LL sum of separate ranks	-255.655	-255.655	-275.441	-275.441
LL with pooled ranks	-261.938	-260.516	-319.496	-307.919
^{2} (nonpooled vs. pooled ranks)[a]	12.57	9.72	88.11	64.96
Pooling of ranks	Accepted	Accepted	Rejected	Rejected

[a] At 1 % significance level, critical values for ^{2} test statistic for 3 and 4 degrees of freedom are 11.34 and 13.28, respectively

Table 3 also reports the results of Hausman-Ruud style rank-ordered heteroscedastic logit models (ROHL). They are obtained by fixing the scale factor for the first rank and estimating relative scale factors for the second and third ranks. The estimates of scale factors are significantly greater than one in the both ROHL models. Being inversely related to the magnitude of random term in the RU model, the magnitude of random term seems to decrease as ranks are added. The consistency of three ranks is strongly rejected with ROHL model for 3 ranks, suggesting that the heterogeneity of responses is not sufficiently systematically related to the ranks for the model to be consistent. Based on these findings, the fixed parameter logit model for three ranks is rejected.

Random Parameter Logit Models

At least two important aspects of modeling strategy must be considered carefully before estimating random parameter models. First, parameters with and without heterogeneity must be selected, preferably by using some prior information. Second, distributions for random coefficients must be specified, typically based on theoretical considerations.

In choosing the heterogeneous and homogeneous coefficients, it is of course possible to allow all the parameters to vary in the population. This strategy relies on the results of the flexibly specified model itself to suggest which parameters are heterogeneous and which not. Following this approach should be supported by prior expectations about parameter heterogeneity. Not only the methodological but also time considerations suggest that. Even with the recent improvements in the computing power, estimating random parameter models can be very time consuming[4]. For instance, the speed of the Train's simulator mainly depends on the number of estimated heterogeneous parameters, and including irrelevant heterogeneous parameters should be avoided. Further, identification is always an issue in estimating random parameters, especially with non-negative/positive coefficients. As it only gets harder with increasing number of random parameters, careful selection of the heterogeneous parameters is recommended.

Random parameters are typically estimated as normally distributed parameters. The normally distributed parameters β_n can get both negative and positive values. They are estimated as $\beta_n=(b_n+\ \ _ne)$, where b and are the estimated mean and deviation parameters of the β_n, and e a standard normal deviate (Train 1998).

Both the theory and common sense often suggest that some random coefficients are non-negatively/positively distributed. In this case, the *BID* coefficient is assumed to be non-positively distributed. For non-positive values of the *BID*, increasing the costs of a policy alternative always decreases its probability to become chosen.

Train (1998) suggests that the non-positive/negative random parameters can be estimated as log-normally distributed, and provides a method for incorporating them into his simulator. Each log-normal β_k can be estimated by expressing them as $\beta_k=exp(b_k+\ \ _ke)$, where b and are estimated mean and deviation parameters of $\ln(\beta_k)$, and e an independent standard normal deviate. Log-normal non-positive parameters are estimated with entering the appropriate exogenous variables as their negative. For the disadvantage of the log-normally distributed random parameters, they are often very hard to estimate and identify (e.g. McFadden and Train 2000).

Alternatively, Layton (2001) proposes employing distributions determined by a single parameter in estimating the non-negative/positive random parameters. While the RP models typically estimate the mean and variance of the RP distribution, the one-parameter distributions (such as Rayleigh-distribution) allow finding all the moments of the RP distribution by estimating just a single parameter. A non-negative parameter *BID* with Rayleigh distribution has a cumulative density function $F(BID)=1-exp[-BID^2/(2b^2)]$ and a probability density function $f(x)=(BID/b^2)exp[-BID^2/(2b^2)]$, where b is the scale parameter fully determining the shape of the distribution. Using the inverse transformation method, the Rayleigh distributed *BID* can be obtained as $BID = (-2b^2ln(1-u))^{1/2}$, where u is a random uniform deviate and b the estimated parameter. The mean, variance, median and mode of the Rayleigh-distributed *BID* are $b(\Pi/2)^{½}$, $(2-\Pi/2)b^2$, $b(log4)$, and b, respectively (Layton 2001).

In the case at hand, both the *BID* and policy alternative dummies were modeled as random parameters. The previous RP applications have typically modeled either the BID or alternative specific dummies as random parameters, not both. With these data, the heterogeneity of preferences for policy alternatives with extensive conservation levels was possible to appear, and random parameter formulation of policy alternative dummies is therefore of specific interest. On the other hand, previous studies suggest that the heterogeneity of preferences is often related to the *BID* coefficient. It was therefore also estimated as a random parameter. Since it essentially represents the negative of the marginal utility of income, it was estimated as a non-positively distributed parameter. Despite continuous and substantial efforts, all the necessary models for

[4] For instance, many of the models reported in this chapter took more than half a day to converge with an up-to-date processor. In addition, several runs with different starting values are often needed to find the global maximum, since the log-likelihood functions of the RP models, unlike their fixed parameter counterparts, are not necessarily globally convex and can therefore have multiple local maximum (McFadden & Train 2000).

this study were impossible to be estimated with the log-normal BID[5]. The BID was therefore expressed as a Rayleigh-distributed random parameter BID – RAYLEIGH[6]. With this specification, convergence was reached much easier and estimation considerably faster.

Table 5 reports the random parameter model results for both the CV and CR data. The pseudo-R^2 of the CR models for 2 and 3 ranks is increased from 0.106 and 0.143 of the fixed logit models to 0.231 and 0.296 of the random parameter models. The explanatory power of the models therefore more than doubled as a result of incorporating unobserved preference heterogeneity.

The CR model for the first rank does not converge with random dummies and its results cannot be reported; a model with fixed dummies and random bid parameter is reported instead. The CR model for first rank performs poorly also in terms of explanatory power. In the CV model, policy alternative dummies are not significant. Explanatory power of the random parameter CV models is substantially higher than for the fixed models; pseudo-R^2 is increased from 0.148 to 0.267.

Table 2.4.4. Random coefficient logit models for CV and CR data

Model:	CV[a]	1 Rank[b]	2 Ranks	3 Ranks
Estimate (*t*-statistic)				
Bid-Rayleigh	1.1309 (3.014)	0.0760 (2.606)	0.1257 (2.396)	0.1578 (3.116)
D25-mean	3.6657 (2.905)	0.3879 (1.393)	1.1133 (3.506)	1.7359 (5.742)
D50-mean	3.5006 (2.882)	0.2965 (0.828)	0.8660 (1.422)	1.4423 (2.313
D75-mean	3.1339 (2.779)	0.1961 (0.458)	-1.8904 (1.332)	-0.2733 (0.281)
D25-dev	-1.6448 (1.304)		0.0122 (0.022)	0.0349 (0.043)
D50-dev	1.5736 (1.290)		3.3713 (4.650)	3.6318 (5.251)
D75-dev	1.7742 (1.417)		6.2289 (4.130)	5.7319 (5.754)
LL	-333.18	-151.11	-224.04	-262.19
LL at zero	-446.78	-162.56	-291.38	-372.66
Pseudo R^2	0.254	0.0704	0.231	0.296

Note: Number of CV and CR observations is 306, and 270, respectively. Simulator with 200 draws were used in estimating the models.

---

5 The estimation of CV models with log-normally distributed *BID* was generally more successful, although excessively time consuming. The estimation of CR models did not succeed. Especially with the models for 2 and 3 ranks, iteration first proceeded seemingly fine for some 80 iterations but then failed to reach the convergence. This could be due to multiple local maxima of the log-likelihood function. However, even the specifications with the *BID* as a sole random parameter did not lead to the convergence.

6 Dave Layton is acknowledged for suggesting this.

[a] Model results for the CV with fixed and random normal dummies are similar and not statistically significantly different. Specification with random normal dummies is chosen for CV data to make results directly comparable to CR, necessary for the data pooling purposes, reported in the next section of the paper.
b Deviations cannot be identified in the model for the first rank, and the results for it are therefore based on model that restricts them as zero.

The estimates of the deviations for the *D50* and *D75* (*D50–dev*, *D75-dev*) are strikingly large and significant, suggesting that the preference heterogeneity for policy alternatives is considerable and should be taken into account while modeling these data. Parameter heterogeneity is also one possible explanation for insignificance of the mean estimates of the *D75*. With highly variable preferences for 75 % policy alternative, estimation cannot provide a significant estimate of the location parameter for distribution of *D75*. Note that the same phenomena was observed in fixed parameter logit models, where significance of *D75* estimate was lower than in the RP models.

The random parameter formulation provided estimation of both CV and CR data with significant improvements. They are next applied together with fixed parameter logit models to combined CV and CR data that estimate pooled models for the CV and CR data.

Pooled models for CV and CR data

Pooled models were estimated using a combined CV and CR data. The estimation can be implemented in several ways; the main concern is to make sure that appropriate likelihood functions are applied to each of the respondents. An indicator variable for the CR data can facilitate estimation. Defining I_{iCR} with a value 1 for CR respondents and a value 0 for CV respondents, the pooled log-likelihood function for individual i is determined as $LL_i = I_{iCR}*P_{iCR} + (1- I_{iCR})*Pi_{iCV}$, where P_{iCR} and P_{iCV} are the appropriate CV and CR response probabilities of the model. Similarly as with the unpooled models, the pooled total log-likelihood function is a sum of the individual likelihoods over the whole sample[7].

Table 6 reports the models results for the combined CV and CR data. A variety of pooled models were estimated to examine the effects of modeling choices on accepting/rejecting pooling the data. The same specification as in unpooled models was applied for pooled models. The unpooled counterparts of all the pooled models can be found from the previous sections of this paper. LR-tests are used for accepting/rejecting the pooling hypothesis; the respective LR-test statistics are reported in the second last row of the Table 6. The test statistics follow ^{2} distribution with degrees of freedom equal to difference in number of estimated parameters between pooled and unpooled models. Estimating a scale parameter in pooled model versions, degrees of freedom for fixed and random parameter logit models with all parameters random equal to 3 and 6, respectively. The respective critical values are 11.34 and 18.48. The LR-test statistic for random parameter models with random BID and fixed policy alternative dummies also has 3 degrees of freedom. If the LR-test statistic is smaller than the critical value, the pooling of data cannot be rejected.

All the results include an estimate of parameter μ_{CR}. It is a scale factor for the CR data, accounting for possible differences in the variance of random term of the RUM between the CV and CR data. As noted before, only the parameter relations are comparable between the different sources of data. Estimating a scale factor allows for direct comparisons of the estimates.

[7] Considerable time savings, especially in estimating random parameter models, can be obtained by structuring the program so that unnecessary calculations are avoided in calculating the log-likelihood function. Calculations of CR response probabilities are uncessary for CV respondents, vice versa.

Logically, if no differences in the random term variance exist between the CV and CR data, the estimate of μ_{CR} is not statistically different from one. Since the scale factor is inversely related to the variance of the random component of the RU-model, an estimate $\mu_{CR} < 1$ suggests that the CR data is noisier than the CV data, and $\mu_{CR} > 1$ the opposite.

Models "CV & 1 Rank" pool the CV model with a CR model for first rank, using both the fixed (FL) and the random coefficient (RCL) formulation. The estimates of the μ_{CR} are statistically significant and smaller than one in both models, suggesting that CR data is noisier. The random parameter model provides a significantly higher explanatory power than the fixed parameter counterpart. Both LR tests statistics for pooling hypothesis are insignificant. Therefore, both fixed and random parameter models provide support for accepting pooling of the CV and CR data.

Table 6. Logit Models for Pooled CV and CR Data (n=576)

Model:	CV & 1 Rank		CV & 2 Rank		CV & 3 Rank	
Estimate (*t*-statistic)	FL	RCL[a]	FL	RCL	FL	RCL
Bid – Fixed	-0.3532 (13.077)		-0.3484 (12.963)		-0.3402 (12.677)	
Bid-Rayleigh		0.8516 (6.351)		1.1286 (2.994)		0.8380 (3.123)
D25 – *mean*	1.5928 (6.669)	2.7152 (7.684)	1.7335 (7.436)	3.7174 (2.876)	1.8513 (7.961)	3.1829 (3.094)
D50 – *mean*	1.3410 (7.120)	2.5373 (8.015)	1.3727 (7.398)	3.5140 (2.853)	1.4679 (7.985)	3.7130 (3.409)
D75 – *mean*	0.9672 (3.697)	2.2661 (5.545)	0.8606 (3.373)	3.1244 (2.757)	0.7131 (2.848)	3.3504 (3.128)
D25 – *dev*				-1.6906 (1.326)		-0.8670 (0.738)
D50 – *dev*				1.6001 (1.297)		2.6625 (2.564)
D75 – *dev*				1.7806 (1.414)		4.2451 (2.925)
μ_{CR}	0.2549 (3.781)	0.0905 (2.902)	0.2617 (5.395)	0.0688 (2.218)	0.2490 (5.996)	0.9103 (3.747)
LL pooled	531.54	485.43	-645.89	613.33	-719.34	-633.91
LL unpooled	531.50	484.29	-642.74	557.22	-700.30	-595.37
LL at zero	609.33	609.33	738.15	738.15	819.432	819.432
LR-test of pooling	0.063	2.28	6.3	112.22	38.08	77.08
Pseudo R^2	0.128	0.203	0.125	0.169	0.122	0.226

Note: Number of CV and CR observations is 306, and 270, respectively. Simulator with 200 draws were used in estimating the models.

[a] Since the CR model for 1 rank with random dummies could not be estimated, a pooled model for CV and 1 Rank CR is also based on model with fixed dummies and Rayleigh bid

"CV & 2 Rank" models pool the fixed and random parameter models for the CV model and the CR model for 2 ranks. Similarly as the previous pooled models, these models result in highly significant estimates with expected signs. Comparing the unpooled and unpooled fixed parameter models results in LR-test statistic 6.3 and accepting the pooling hypothesis. However, the pooled random parameter model strongly rejects pooling of the CV and CR data. Despite rejection of pooling, the random parameter model results in substantially higher pseudo-R^2 than the fixed parameter model.

Models "CV & 3 Rank" pool fixed and random parameter model for 3 ranks and the CV data. Results from these models are similar with the pooled models with CR model for two ranks. However, pooling of CV and CR data is now strongly rejected for both the fixed and the random parameter models. Rejecting the pooling with the random parameter models, together with the results of pooled models with CR model for 2 ranks, suggest that differences in parameter heterogeneity between the two data are a possible source of their inequality.

The overall similarity of the parameter estimates across all the models is distinctive for the pooled model results. This is likely to have resulted from more precise estimates of the CV models data "dominating" the identification of estimates for the pooled models. With lower variability of the random term of RUM, as suggested uniformly smaller than one estimates of the μ_{CR}, the CV data plays a relatively more important role in identifying the estimates, even with almost equal number of CV and CR observations.

Although not reported in Table 6, the pooling hypothesis was further tested with the restriction $\mu_{CR}=1$ that imposes equal variances of the RUM random terms for the CV and CV models. Using the completely pooled model, the pooling of the CV and CR data is rejected using all the models. The pooled fixed model for CV and first rank CR data with a CR scale factor provides the strongest support for accepting pooling, and is therefore the likeliest candidate to provide support for the complete invariance hypothesis. Estimating the pooled model for it results in a value 45.5 for the LR test statistic, strongly rejecting the complete pooling of the CV and CR data. Other models are less likely to provide support for complete pooling hypothesis, and complete invariance of CV and CR data is therefore uniformly rejected.

Willingness to Pay Estimates

Including policy implementation costs and policy specific dummies in estimated models allows capturing WTP for different policy scenarios indirectly from the results. The mean WTP for policy alternative x_j is calculated as (e.g. Goett et al. 2000)

$$\frac{\partial U / \partial x_j}{\partial U / \partial y} \tag{2.21}$$

The $\partial U / \partial x_j$ is measured by the alternative specific dummies *D25*, *D50* and *D75*, and the $\partial U / \partial y$ by the *BID* estimate. The mean WTP estimates for the fixed logit estimates are calculated as *D25/BID*, *D50/BID* and *D75/BID*. The means of normally distributed random parameters equal their estimates and calculation of WTP is similar as with the fixed parameter models. The means for Rayleigh distributed *BID* must be calculated as described earlier in this paper.

Table 7 reports the mean WTP estimates for the estimated models. The results are divided into fixed and random parameter models for unpooled and pooled data. The estimates for the CR 1 rank model were not statistically significant, and the WTP estimates are therefore not presented for them. In the previous analyses, the following models were clearly rejected because of inconsistency of rankings or failure to accept data pooling hypothesis: (a) Fixed logit model "CR-3 rank data", (b) Pooled fixed logit model "CV & CR – 3 rank", (c) Pooled random parameter model "CV & CR-2 rank" and (d) Pooled random parameter model "CV & CR-3 rank". Their results are expressed in *(italics)*.

Table 2.4.6. Willingness to pay estimates (US$) [a, b,c]

Policy alternative	Fixed Parameter Logit			Random Parameter Logit		
Model type	25 %	50%	75 %	25 %	50 %	75 %
CV	**73**	**61**	**44**	**42**	**40**	**36**
CR – 2 Ranks	**156**	**115**	*insign.*	**95**	**92**	**53**
CR – 3 Ranks	*(315)*	*(271)*	*insign.*	**142**	**118**	*insign.*
Pooled models						
CV + CR – 1 Rank	**73**	**61**	**44**	**41**	**38**	**34**
CV + CR – 2 Rank	**80**	**64**	**40**	*(42)*	*(40)*	*(34)*
CV + CR – 3 Rank	*(88)*	*(70)*	*(34)*	*(49)*	*(57)*	*(51)*

[a] Insignificant estimates are expressed with *insig.*
[b] US$=FIM 6.2.
[c] WTP estimates for the rejected models are reported in *italics* in brackets.

Results for the 75 % conservation alternative are very similar across all the models providing it with a significant estimate. The WTP for it varies between US$ 36 and 53. Estimates for the mean WTP for 50 % alternative cover a considerably wider range from US$ 38 and 118, similarly as the mean WTP for 25 % conservation alternative with values between US$ 41 and 156 .

All the significant and accepted WTP models result consistently higher WTP for the 25 % alternative than for the 50 % alternative, and in turn for higher WTP for the 50 % alternative than for the 75 % alternative.

Figure 1 graphs the WTP estimates. Each set of bars graphs estimates for one set of models. The first set "CV" stands for all the CV models; its first three bars in the left are the fixed model results for 25%, 50%, and 75 % programs. Three bars on the right are the estimates based on the random parameter models. The "CR1" has no bars since none of the estimated WTP figures were statistically significant. The "CR2" and "CR3" graph the WTP estimates for the CR models for 2 and 3 ranks. Note that the fixed 3 rank model for the CR data was rejected in testing for the consistency of rankings. "CV+CR1", "CV+CR2", and "CV+CR3" graph the WTP estimates of the pooled models. Note also that the random parameter model for the pooled CV and CR data on 2 ranks was rejected. Further, both the fixed and random parameter models for pooled CV and 3-rank CR data were rejected.

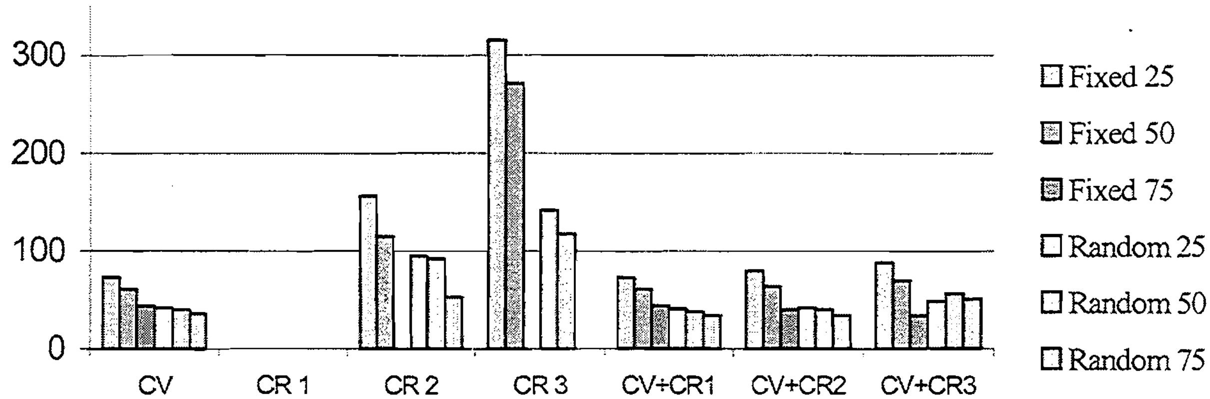


Figure 1. Estimates of the WTP for fixed and random parameter models for the CV and CR data

The results consists of variety of WTP estimates based on different models. The question becomes which results are preferred and chosen for further purposes, such as policy evaluation. Clearly, all the rejected models can be screened out first. The rest of the models can be evaluated by using success in pooling the CV and CR data as criteria. The models that successfully pool the CV and CR data are not fully contingent on a single survey method, and can therefore considered most general.

Using the pooling success criteria leaves us with a choice from the following three models: (1) Pooled fixed parameter model for the CV data and the CR data on 1 rank, (2) Pooled random parameter model for the CV data and the CR data on 1 rank, and (3) Pooled fixed parameter model for the CV data and the CR data on 2 ranks. The fixed model with 1 rank CR data can be screened out as more restrictive than the random parameter model with 1 rank, and as less detailed than the fixed parameter model with 2 ranks. Both remaining models have certain advantages. The random parameter model with 1 rank is less restrictive than the fixed model with 2 ranks, but the fixed parameter model with 2 ranks utilizes the data in more detail than the model with 1 rank. However, the random parameter models statistically outperformed the fixed models throughout the analysis, and the pooled random parameter model for the CV data and the CR data on 1 rank is therefore chosen as the preferred approach for modeling these data.

Discussion

This study examined different econometric modeling strategies for CV and CR survey data. Both conventional fixed logit models and recently developed random parameter logit models were reviewed and applied to the data at hand. The results provided another confirmation that considerable care must be practiced in applying fixed parameter logit models. Especially the fixed parameter models for the CR data on full rankings of four alternatives violated assumptions of the conditional logit model.

Applying data pooling techniques in testing for equality between CV and CR was another objective of this study. Successful pooling of the CV and CR data required estimation of scale factors for separate data sources, which then made the parameter estimates comparable between separate sources of data. Without the scale parameter, pooling was uniformly rejected. With scale factors in the estimated model, pooling of the CV and CR data could not be uniformly rejected or accepted. The more detailed models for the CV and CR responses are likely to reject the pooling hypothesis. Less detailed models such as the CV and conditional choice logit models generally did not provide sufficient statistical evidence for rejecting the pooling hypothesis.

The random parameter models do not seem to provide a miracle in terms of solving differences between the CV and CR data, although there is some evidence that fixed logit models could exaggerate the differences between the WTP results for the CV and CR data. Especially the results of the rejected fixed logit models for the three ranks CR data suggest that. This could be related to the model specification, since the WTP results are typically sensitive to the chosen specification. Examining different specifications is one of the objectives for future research.

Data pooling techniques provide a powerful approach to tests for invariance between the different sources of data. The analysis of pooled data that has been presented in this paper highlights only some of the possible uses of data pooling methods in examining SP survey methods. For instance, sources of differences between different SP sources can be further examined using the same framework.

Issue of negative marginal WTP for conservation after reaching a certain conservation levels, is of course very interesting and worth some further investigation. Given the opposition of some of the public for increasing the forest conservation in Finland, the result is not necessarily a surprising finding. Further, a major conservation program called Natura 2000 was under preparation around the time of the survey. The Natura 2000 is part of the European network of conservation areas, which in turn is the core policy for nature conservation in the EU. Preparation of the Natura 2000 was carried out without public hearings until the first official proposal was released. The program covers all the existing conservation areas, but also introduces a large number of new areas with various restrictions in their use. Many of the new areas are on private lands, and releasing the first proposal for the Natura 2000 resulted a public outcry, including literally thousands of appeals. The public opinion also criticized the way in which the program was prepared. Therefore, it is a logical finding that most Finns support moderate increases in conservation, but are reluctant to show support to extensive conservation programs, that they may view a regulatory "overshooting".

References

Adamoviz, W., Louviere, J. & Williams, M. 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26:271-292.

Amemiya, T. 1985. *Advanced Econometrics*. Harvard University Press. 521 p.

Beggs, S., Cardell, S. & Hausman, J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics* 16:1-19.

Boxall, P., Adamoviz, W., Swait, J., Williams, M. & Louviere, J.1996. A comparison of stated preference methods for environmental valuation. *Ecological Economics* 18:243-253.

Brownstone, D. & Train, K. 1999. Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics*, Vol. 89(1):109-129.

Chapman, R.G. & Staelin, R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* 19: 288-301.

Desvouges, W.H. & Smith, V.K. 1983. Option price estimates for water quality improvements: a contingent valuation study for Monogahela river. *Journal of Environmental Economics and Management* 14:248-267.

Goett, A., Hudson, K. & Train, K. 2000. Customers' Choice Among Retail Energy Suppliers: The Willigness-to-Pay for Service Attributes. Department of Economics, University of California, Berkeley.

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

Hanemann, M. W., Loomis, J. & Kanninen, B. 1991. Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics* 1255:1263.

Hanemann, M.W. & Kanninen, B. The Statistical Analysis of Discrete-Response CV Data. In Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation (ed. Bateman, I & Willis, K.).

Hausman, J.A. & Ruud, P.A. 1987. Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics* 34:83-104.

Hensher, D, Louviere, J.J. & Swait, J. 1999. Combining sources of preference data. *Journal of Econometrics* 89:197-221.

Layton, D.F. 2000. Random Coefficient Models for Stated Preference Surveys. *Journal of Environmental Economics and Management* 40:21-36.

Layton, D.F. 2001. Notes on One Parameter Family of Distributions for Random Coefficient Models. Department of Environmental Policy and Science. University of California Davis.

Louviere, J.J. & Hensher, D.A. 1983. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research* 10:348-361.

Magat, W.A., Viscusi, W.K., Huber, J. 1988. Paired comparison and contingent valuation approaches to morbidity risk valuation. *Journal of Environmental Economics and Management* 15:395-411.

McFadden, D.1974. Conditional logit analysis of qualitative choice behavior, in Zarembka, P. (ed.), Frontiers in Econometrics, Academic Press, New York.

McFadden, D. & Train, K. 2000. Mixed MNL Models for Discrete Response. Forthcoming in *Journal of Applied Econometrics.*

Revelt, D. & Train, K. 1998. Mixed Logit with Repeated Choice: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics.* Vol. 80: 647-657.

Roe, B., Boyle, K.,J. & Teisl, M.,F. Using Conjoint Analysis to Derive Estimates of Compensating Variation. *Journal of Environmental Economics and Management* 31:145-159.

Stevens, T.H., Belkner, R., Dennis, D., Kittredge, D. & Willis, C. 2000. Comparison of contingent valuation and conjoint analysis in ecosystem management. *Ecological Economics* 32: 63-74.

Swait, J. & Louviere, J.J. 1993. The role of scale parameter in the estimation and use of generalized extreme utility models. *Journal of Marketing Research* 10:23-45.

Train, K. 1998. Recreation Demand Models with Taste Differences Over People. *Land Economics* 74(2):230-39.

Estimating the Economic Value of Big Game Habitat Production from Natural and Prescribed Fire

John Loomis
Department of Agricultural and Resource Economics
Colorado State University

Dana Griffin
Department of Agricultural and Resource Economics
Colorado State University

Ellen Wu
Department of Agricultural and Resource Economics
Colorado State University

and

Armando Gonzalez-Caban,
Forest Fire Lab
Pacific Southwest Station, USDA Forest Service

Presented at the Annual Meeting of Western Regional Project W133
Miami, Florida
February 26-28, 2001

Acknowledgements: I wish to thank Dan McCollum and Diane Freeman, USDA Forest Service and Jane McKeever of the California Department of Fish and Game for assistance in the study design and data collection. I would like to thank Lucas Bair of Colorado State University for research assistance regarding the cost of prescribed burning and editorial assistance.

Abstract: This paper quantifies the deer harvest and deer hunting benefits of prescribed burning on the San Jacinto Ranger District of the San Bernadino National Forest in Southern California. Two approaches to estimating a production function relating deer harvest response to prescribed burning were compared. The San Jacinto Ranger District time series model showed a statistically significant effect of prescribed and wildfire on deer harvest over the 20-year period of 1979-1998. The micro GIS model specifications showed the initial effect of prescribed burning had a statistically significant effect of prescribed burning on deer harvest in the 37 hunting areas within the San Jacinto Ranger District. Lagged effects of prescribed burning were consistently insignificant in our models. The macro time-series model estimated a larger response to burning of the first 1,000 acres than the micro GIS model did, but for increases in fire beyond 1,000 acres, the two models provide nearly identical estimates. The net economic value of the resulting additional deer hunting benefit was estimated using the Travel Cost Method (TCM) at $257 per additional deer harvested due to the additional trips the hunter took in response to increasing deer harvest. While the initial deer hunting benefit response to the current magnitude of prescribed burning of 1,100 acres ranges from $4,112 to $8,481 depending on the model, the incremental gains for additional prescribed burning are quite similar across models. The costs of prescribed burning greatly exceed these benefits, suggesting that prescribed burning should is economically justified only if it provides substantial wildfire hazard reduction benefits and other multiple use benefits beyond deer hunting.

Introduction

This research compares two different approaches to evaluating the effectiveness of prescribed burning for increasing big game wildlife habitat as well as providing results for the San Jacinto Ranger District (SJRD) in the San Bernadino National Forest located in southern California. The methodological contribution begins to answer the challenge posed by Hesseln (2000) in her recent review of the economics of prescribed burning. She stated "... however, there is a lack of economic models to evaluate short- and long-term ecological benefits of prescribed fire. Without understanding the relationship between economic outcomes and ecological effects, it will be difficult to make effective investment decisions. Research should focus on defining a production function to identify long-term relationships between prescribed burning and ecological effects. Identifying production functions relationships will form the basis for future cost-benefit analysis with respect to prescribed burning." (Hessln, 2000: 331-332). This is a similar challenge faced by the Clearwater National Forest Elk Initiative which is considering using prescribed burning to increase elk populations in Idaho. This study demonstrates two different approaches to estimating production relationships between prescribed burning and deer harvest using time series data and Geographic Information System (GIS) approaches to make a first modest step in the direction suggested by Hesslen. The paper provides a template for evaluating other wildlife prescribed burning programs such as the Clearwater National Forest Elk Initiative.

Background on the Study Area

The San Jacinto Ranger District (SJRD) is located in Southern California's San Bernardino National Forest near Palm Springs. As noted by the USDA Forest Service, "Some of the best deer hunting in Riverside County is found in this area." (Gibbs, et al., 1995: 6). The SJRD is an ideal area to demonstrate and compare different approaches to estimating a production function between prescribed burning and deer harvest because prescribed fire has been used for more than 20 years to stem the long-term decline in deer populations since the 1970's (Paulek, 1989, Gibbs, et al., 1995). Previous research on prescribed burning shows that fire enhances deer habitat and populations (DFG, 1998) but the economic benefits have not been quantified. The results of our analysis should be of some policy relevance as the San Jacinto Ranger District plans to increase the amount of prescribed burning by 50% to 100% over the next few years (Walker, 2001; Gibbs, et al., 1995).

In general, Southern California is characterized by a Mediterranean climate, with hot and dry summers and cool, moist winters. There is a significant amount of variation in temperatures and local site conditions in the Ranger District in the San Jacinto and Santa Rosa Mountains. Elevations in these ranges reach 10,800 feet. The dominant vegetation within the SJRD below 5000 feet is chaparral. Annual rainfall for the chaparral biome is approximately 15 to 16 inches. Areas above 5,000 feet tend to be dominated by hardwoods and conifers such as live oak and Douglas fir with annual rainfall reaching up to 30 inches.

Within the San Jacinto Ranger District, the land is primarily managed by the USDA Forest Service, with small amounts of land administered the State of California as the Mount San Jacinto State Park. The land within the San Jacinto Ranger District is an area that evolved with fire as a natural environmental factor. Declining abundance of successional vegetation communities is considered to have the greatest long-term effects on deer populations (CDFG, 1998). Historically, fire, either prescribed or natural, has been the primary mechanism for establishing these vegetation communities. Studies in California have noted that after a burn, increased deer numbers can be attributed to individuals moving into the area to feed (Klinger et al.1989). These increased deer numbers have been thought to improve reproduction due to increased forage quality and an increase in fawn survival rates. The California Department of Fish and Game has noted a significant increase in buck harvest from 1987 to 1996 in hunt zones that had large fires, versus hunt zones that did not have large fires (CDFG, 1998). To improve deer habitat in California, controlled burning has been underway in all the major parks and forests for many years (Kie, 1984). Efforts including controlled burning to remove brush have been part of a program to create desirable deer habitat to mitigate the loss of deer habitat resulting from commercial and residential development.

Two Production Function Modeling Approaches

The purpose of this study is to test whether prescribed burning has a systematic effect on deer harvest. By examining prescribed burning on deer harvest with two different approaches, a macro or aggregate time series approach and a micro, spatial approach (e.g., GIS), then comparisons can be made between the results for consistency between these two approaches. A macro approach would be able to test the effects of fire, prescribed and natural, across the entire study area over a long period of time. Using a micro approach provides greater spatial detail, such as the influence of a meadow or ridge but less temporal time frame is covered due to data limitations. Thus each approach to estimating the production function has its relative strengths and weaknesses and each method may be more useful than the other depending on the data availability.

Estimating a production function that relates deer harvest to acres of prescribed burning must also control for other inputs that influence the production of deer for harvest. This includes wildfire, elevation (used as a proxy for vegetation data which was incomplete), rainfall, temperature and distance to roads. Thus, multiple regression is an appropriate technique.

Time Series, Macro Scale Production Function

The first approach is based on a time series regression model to test for a relationship between deer harvest in the SJRD and prescribed fire, controlling for other independent variables such as annual precipitation and temperature during the hunting season. This approach used a dataset for SJRD, provided by the California Department of Fish and Game (CDFG) and the USDA Forest Service. The fire records provided data from 1979 for wildfire and prescribed burns within the San Jacinto Ranger District. This ranger district represents the majority of publicly accessible land for deer hunting in Riverside County. Deer harvest data from 1979, was provided by CDFG. The full model is given as equation (1.1), and then a lagged model is included as equation (1.2), which allows for harvest to be sensitive to previous years prescribed fire and wildfire. In past research the use of burned areas by deer increases dramatically during the following years (Klinger, 1989). Therefore, this model tests for these effects the following years by using a lag on the fire variables.

SJRD Time Series Production Function Model:

(1.1) SJRD Deer harvest in year $_t$ = func ($RXFIRE_t$, $WILDFIRE_t$, $TOTPRECIP_t$, OCTTEMPt, $YEAR_t$)

Where: $RXFIRE_t$ = the acres of prescribed fire in year t
$WILDFIRE_t$ = the acres of wildfire in year t
$TOTPRECIP_t$ = the sum of precipitation for year t
$OCTTEMP_t$ = temperature in October during the hunting season
$YEAR_t$ = a trend variable, with 1979 = 1, 1976 = 2, ... 1998= 19

A non-linear form of equation 1.1 is estimated using the log-log form. This format allows for a non-linear relationship and the coefficients for fire can be interpreted as elasticity's. This is the percent change in deer harvest with a 1% change in acres burned.

When modeling the aggregate harvest for all of the San Jacinto Ranger District, the dependent variable is the total number of deer harvested in year t. This is a relatively large number and varies between 80 to 157 deer in any given year. Therefore using Ordinary Least Squares (OLS) is an acceptable approach for the macro time-series modeling.

Micro GIS Approach to Estimating the Production Function

The second approach taken in this study focuses on using a geographic information system (GIS) for integrating spatial data into an economic relationship. A similar multiple regression approach was used as in the first method, except now the study area was divided into 37 individual hunting zones delineated by California Department of Fish and Game (CDFG) instead of treating the entire SJRD as one unit. These hunt areas are defined by topographic features such as steep ridgelines or sometimes human features such as towns or major roads. This allowed for the incorporation of other influences on deer harvest that varied spatially across individual hunting areas such as distance to roads and elevation.

The first step in the GIS analysis was to identify the necessary layers needed to run a regression between deer harvest and fires. The following layers were constructed for the regression model: a harvest layer, which contains deer harvest by hunt zones, which serves as the dependent variable. Then layers were added for the independent variables including acres of prescribed burning, wildfire acres, average elevation, temperature, distance to trails, dirt roads and roads from each hunting zone, and distance to wildfires from each hunting zone. Vegetation type would have been desirable, but this information was incomplete and will not be completed for the entire area until well into the future. The models developed for the harvest areas accounted for the non-uniform size of each hunting zone using one of two approaches: (1) include the size of the harvest area as a separate independent variable and use total acres of an area burned and, (2) transform the dependent variable into deer harvest per acre, then use an OLS regression. The total area model is shown in equation (1.2):

(1.2) Model with harvest a function of total size of fire, including lags

Deer harvest in $area_i$ in $year_t$ = func (Avg_Elev_i, $Ltotal\_Wildfire_{i\ t}$, $Ltotal\_Wildfire_{i\ t-1}$, $Ltotal\_Wildfire_{i\ t-2}$, $Ltotal\_Wildfire_{i\ t-3}$, $Ltotal\_Rxfire_{i\ t}$, $Ltotal\_Rxfire_{i\ t-1}$, $Ltotal\_Rxfire_{i\ t-2}$, $Ltotal\_Rxfire_{i\ t-3}$, $Ldirt\_distance_i$, $Ltrail\_distance_i$, $LHvst\_Area_i$, Oct_Temp_t, $Year_t$)

The model in equation (1.2) was estimated using count data models instead of OLS regression. The reason is that at the micro level where harvest in any limited spatial unit is a small non-negative integer variable, count data model models are statistically more efficient because such models are based on probability distributions that have mass only at nonnegative integers (Hellerstein, 1992). This is certainly the case for deer harvests as hunters cannot harvest a fraction of a deer and the number harvested in each unit is typically 0,1,2,3... rather than 10 or 50. One of the simplest count distributions is the Poisson process. Given the stringency of the mean-variance equality restriction imposed by the Poisson, a more generalized count model like the negative binomial, is often more consistent with the data. The negative binomial version relaxes the mean-variance equality of the Poisson. Both the Poisson and the negative binomial yield the equivalent of a semi-log form where the log of the dependent variable is regressed against the explanatory variables.

An alternative specification to account for the different size harvest areas, involved transforming the dependent variable into deer harvest per acre. This results in equation (1.3), which is estimated using ordinary least squares regression since this dependent variable is continuous and does not have to take on integer values:

(1.3) Model based on Deer Harvest per acre, using OLS, Log-Log Form:

Log Deer_harvest per acre in $\text{year}_{i\ t}$ = func (Avg_Elev_i, $\text{Ltotal\_Wildfire}_{i\ t}$, $\text{Ltotal\_Wildfire}_{i\ t-1}$, $\text{Ltotal\_Wildfire}_{i\ t-2}$, $\text{Ltotal\_Wildfire}_{i\ t-3}$, $\text{Ltotal\_Rxfire}_{i\ t}$, $\text{Ltotal\_Rxfire}_{i\ t-1}$, $\text{Ltotal\_Rxfire}_{i\ t-2}$, $\text{Ltotal\_Rxfire}_{i\ t-3}$, Ldirt_distance_i, Ltrail_distance_i, LHvst_Area_i, Oct_Temp_t, Year_t)

Description of GIS Based Micro Regression Variables

Elevations are based on USGS digital elevation models (DEM's) and act as a proxy for vegetation types, which were not available. However, we do not have an expected sign on elevation, but simply wish to control for elevation differences between the 37 individual hunting areas within the San Jacinto Ranger District. Both fire variables, wildfire and prescribed fire, are expected to have a positive sign. This expectation is based on suggestions of past literature. The distance to road and trail variables are based on the distances from a central point in each hunting zone. Two arguments can be made about the variables direction, therefore the expectation is left to be ambiguous. One argument is based on accessibility for hunters, where having a close proximity to either a trail or road would make hunting easier and more desirable, which would positively effect deer harvest. The second argument is based on the intrusion of deer habitat by a road or trail. This perspective would lead to a decline in deer harvest because roads cause a break in habitat and pose as a threat.

The distance to fire variable is based on distance from a central point in each hunting zone to the closest fire in that time period. This variable's sign may be either positive or negative.

Harvest area, which takes into account the size of each hunting zone, is expected to have a positive sign. The argument here is that as hunting areas become larger, then the amount of deer habitat increases, which attracts more deer, therefore the probability of hunter success increases. October temperature and year are the other variables used in the models. October is when hunting season is open, and based on hunter's surveys, when temperatures are high deer tend to bed down and seek cover. Therefore, harvest rates decline which gives the October temperature a negative sign. Year is a trend variable to capture any temporally varying effects and does not carry any expected sign. Table 1 summarizes the description of the variables and their expected sign, if any.

Table 1 Description of GIS Based Micro Regression Variables

Variable	Description	Expected sign on Coefficient
Deer harvest	The dependent variable, deer harvest is the number of deer harvested in a designated hunting zone.	
Avg_Elev	Average Elevations, based on USGS Digital Elev. Model and re-classed into elevation categories.	No expectation
Total_Wildfire	Total Wildfires in a particular year within the San Jacinto Ranger District	+
Total_Rxfire	Total Prescribed Fires within the San Jacinto Ranger District for a particular year.	+
dirt_distance	The distance to the nearest dirt road, in meters from a central location of each hunting zone	No expectation
trail_distance	The distance to the nearest trail, in meters from a central location of each hunting zone.	No expectation
Fire_dist	The average distance from a central location of each hunting zone to the central point of a wildfire.	No expectation
Hvstarea	The size of each harvest area, measured in acres.	+
Oct_Temp	The average temperature in October, degrees Fahrenheit	_
Year	A trend variable to look for systematic changes	No Expectation

Estimated Production Functions

Macro- Time Series San Jacinto Ranger District Equations

Allowing for non-linearity proved to be a better predictor of deer harvest than the linear models (linear results available from the authors) so a double log model is presented. Results from preliminary regressions also suggested combining the wildfire and prescribed burn into one variable. In Table 2, the coefficient for total fire has a small magnitude of .048, but it has a significant t-statistic of 2.3. The sign on this variable is positive and the coefficient can be interpreted as elasticity's using the log-log form. Therefore, a one percent increase in acres burned will lead to a .048% increase in deer harvest. The other significant variable is October temperature and year, a trend variable. Again a negative sign on the October coefficient, relates to observations that an increase in temperature results in a decrease in the number of deer harvested. The year variable indicates that a systematic affect exists within the model. This models explanatory power is reasonably good with an R^2 value of .67. The Durbin-Watson statistic of 2.06 indicates that autocorrelation is not a problem. The same model indicated in Table 2 was estimated with a 1-year lag but this model did not perform well using the lag. The coefficient on the lag of total fire (-1) was .01 and the t-statistic is .44, which indicates the lag is insignificant. The R^2 value did not change from the previous model (results available from authors).

Table 2 Macro Time Series Ranger District Log-Log Model

Dependent Variable: SJRD HARVEST N=19				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	41.80865	11.0701	3.776719	0.002
ln_Totalfire	0.048735	0.020547	2.371872	0.0326
Total_Precip	-0.00096	0.002621	-0.366608	0.7194
Oct_Temp	-0.02703	0.010659	-2.536153	0.0237
Year	-0.01785	0.00558	-3.199292	0.0064
R-squared	0.67722	Mean dependent var		4.808914
Adj R-squared	0.584997	S.D. dependent var		0.202321
S.E. of regression	0.130337	F-statistic		6.343287
Durbin-Watson	2.066171	Prob (F-statistic)		0.002096

Summary of Micro Regressions Based on GIS Analysis

The two regression models estimated using GIS derived data are presented in this section: one count data and the other OLS, both of which show prescribed burning had a statistically significant effect on deer harvest. As can be seen in Table 3a, total acres of prescribed fire is significant during the year of the prescribed fire, and its significance declines over the next three years in the count data model. During the first year prescribed fires coefficient is .044 with a t-statistic of 2.4. Since this count data model logs the fire acreage variables it is equivalent to a log-log model. As such, the .044 is the elasticity, which is remarkably similar to the .048 elasticity in the macro time-series model reported in Table 2. Total acres of wildfire was not significant for any of the years in this equation. The total area count data model has an R^2 value of .25.

Using OLS as an estimator of deer harvest per acre as a function of fire and the other variables provides a similar pattern of signs and significance as the total area count data equation. In this model, a double log form was also used, but this time the dependent variable acts as a controlling measure for the size of each harvest area by dividing harvest in each hunting zone by the number of acres in each zone. The result of this model in Table 3b shows that prescribed burning has a statistically significant effect on deer harvest in the first year with a t-statistic of 2.25. Then during the years following the fire, prescribed burning becomes less significant, which corresponds to the previous count data model. The only time wildfire has a significant impact is during the second year following the burn. The sign of the coefficient for wildfire in the second year is negative and less than one, which would imply a negative effect on deer harvest in that year. Distance to dirt roads is also significant, a t-statistic of 5.17 and a negative coefficient -.012. This means that harvest areas further away from dirt roads have a lower probability of harvesting a deer. The distance to trails variable implies having a distant proximity to trails increases the probability of a deer harvest. All the other variables in this model fail to be significant indicators of deer harvest, except for the trend variable, year. Therefore, some unidentifiable systematic temporal change is occurring within the model. Overall this model has a lower level of explanatory power than the total area micro count data model. The R^2 value using OLS is .13 as compared to twice this level of explanatory power in the total area count data model.

Table 3a. Count Data Model Based on GIS using Total Acres Burned with Lags

Dependent Variable: DEERKILL; n=825
Method: ML - Negative Binomial Count

	Coefficient	Std. Error	t-Statistic	Prob.
C	62.96425	23.11578	2.723865	0.0065
LAVG_ELEV	-0.237276	0.130702	-1.815396	0.0695
LTOTWFIRES	0.010712	0.017143	0.624876	0.5321
LTOTWFIRES(-1)	0.008299	0.017016	0.487728	0.6257
LTOTWFIRES(-2)	-0.027728	0.015488	-1.790263	0.0734
LTOTWFIRES(-3)	-0.02466	0.015578	-1.582984	0.1134
LTOTRXFIRES	0.044067	0.017907	2.460872	0.0139
LTOTRXFIRES(-1)	0.027531	0.02701	1.019289	0.3081
LTOTRXFIRES(-2)	0.011491	0.02223	0.516926	0.6052
LTOTRXFIRES(-3)	0.011491	0.018667	0.615537	0.5382
LDIRTDIST	-0.233799	0.037744	-6.194352	0
LTRAILDIST	0.395161	0.041757	9.463323	0
LFIRE_DIST	0.072684	0.047397	1.533514	0.1251
LHUNTAREA	0.940678	0.086997	10.81281	0
OCT_TEMP	-0.012073	0.016818	-0.717869	0.4728
YEAR	-0.034733	0.01176	-2.953497	0.0031
	Overdispersion Parameter			
Alpha: C(17)	-0.281026	0.108144	-2.598621	0.0094
R-squared	0.256944	Mean dependent var		1.758788
Adjusted R-squared	0.24223	S.D. dependent var		2.611355
S.E. of regression	2.273183	Avg. log likelihood		-1.617564
Restr. log likelihood	-1920.633	LR index (Pseudo-R2)		0.305182

3b. Least Squares Deer Harvest per Acre using GIS Data Model with Lags

Dependent Variable: LDEERKILLAC; n=825
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.341834	1.588264	0.844843	0.3984
LAVG_ELEV	-0.009703	0.009275	-1.046114	0.2958
LTOTWFIRES	0.001163	0.001208	0.963236	0.3357
LTOTWFIRES(-1)	0.000499	0.001164	0.429264	0.6678
LTOTWFIRES(-2)	-0.002231	0.00107	-2.08622	0.0373
LTOTWFIRES(-3)	-0.001775	0.001091	-1.627588	0.104
LTOTRXFIRES	0.002635	0.001169	2.254844	0.0244
LTOTRXFIRES(-1)	0.002134	0.00188	1.134854	0.2568
LTOTRXFIRES(-2)	0.001333	0.001547	0.862046	0.3889
LTOTRXFIRES(-3)	0.001212	0.001308	0.926172	0.3546
LDIRTDIST	-0.012952	0.002503	-5.174828	0
LTRAILDIST	0.01825	0.002222	8.213173	0
LFIRE_DIST	0.005144	0.003201	1.607198	0.1084
LHUNTAREA	-0.008678	0.006156	-1.409754	0.159
LOCT_TEMP	-0.050413	0.086045	-0.585888	0.5581
YEAR	-0.002827	0.000823	-3.436417	0.0006
R-squared	0.138688	Mean dependent var		-4.532693
Adjusted R-squared	0.122718	S.D. dependent var		0.093204
S.E. of regression	0.087298	F-statistic		8.684311
Durbin-Watson stat	1.266619	Prob (F-statistic)		.0000

Applying the Regression Production Functions

To calculate the incremental effects of different levels of prescribed burning on deer harvest, the acres burned variable is increased from one level to a higher level in the regression model. We use the double-log macro time-series model and the micro GIS based double-log total area count data models, as these two models have the highest explanatory power. The resulting predicted change in deer harvest will be valued in dollar terms in the next section.

Applying Results of Micro GIS Production Function Model

The results of the "Total Acres Burned" count data model from Table 3a provides positive evidence on the desirable effects of prescribed burning programs on deer harvest. The first row in Table 4 forecasts the estimated number of deer that would be harvested without having a prescribed burning program. The effect of further increasing prescribed burning is then calculated by increasing the number of acres burned in each of the 37 hunting areas by 100 acres and then 200 acres to provide a wide range of prescribed burning levels in the SJRD. The first level (1,100 acres) is about the average prescribed burning over the last 20 years. Maintaining this level of prescribed burning does provide a significant increase in deer harvest over the no burning level. However, the gain in deer harvest increases more slowly with additional increases in burning in each hunt area.

Applying Results of Macro Time Series Production Function Model

To estimate the change in deer harvest using the Macro Time Series Production Function Model, the double log model reported in Table 2 is used. The total fire variable in this model is increased and the predicted level of deer harvest is calculated at the mean of the other variables. This is done at the same four acreage levels used above.

Table 4 Comparison of Deer Harvest Response to Prescribed Burning Using the Macro Time Series Model and GIS Micro Model.

RX acres burned	Additional acres Burned	Macro Time Series Model: # Deer Harvested	Time Series Marginal increase in Deer Harvest	GIS Micro Model: # Deer Harvested	GIS Marginal increase in Harvest
1	NA	83	NA	42	NA
1100	1110	116	33	58	16
4810	3700	124	8	66	8
8510	3700	128	4	71	5

The results in Table 4 suggest there is a substantial gain in deer harvest with the first 1,100 acres burned, especially as calculated from the macro time-series model. However a very similar diminishing marginal effect is evident from both the macro time-series production function regression and the micro GIS production function regression after burning more than 1,100 acres. That is, regardless of the spatial level of detail adopted, burning an additional 3,700 acres is expected to result in about eight more deer being harvested in the SJRD.

In order to determine the economic efficiency of additional prescribed burning it is necessary to compare the benefits of additional prescribed burning in the form of the economic value of deer harvest against the costs. It is to the development of the valuation data that we now turn.

Valuation of Deer Hunting

According to CDFG, deer hunting is considered as one of the major outdoor recreation activities in SJRD year. Previous research on deer hunting in California showed that increased success rates and opportunities to harvest a trophy deer increase the economic value of deer hunting (Loomis, et al. 1989, Creel and Loomis 1992).

Travel Cost Method for Valuation of Deer Hunting

The Travel Cost Method (TCM) has been a primary approach for valuing recreational hunting. The basic concept of TCM is the travel cost (i.e. transportation cost, travel time) to the site is used as the proxy for the price of access to the site. When hunters are surveyed and asked questions about the number of trips they take and their travel cost to the site, enough information can be generated to estimate a demand curve. From the demand curve, net willingness to pay or consumer surplus can be calculated.

Besides variable travel cost or its proxy, travel distance, many articles discuss the inclusion of a travel time variable in the demand function. Knetsch (1963) was the first to point out the opportunity costs of time is part of travel costs as well. Cesario (1976) suggested one-fourth the wage rate as an appropriate estimate of the opportunity cost of time based on commuting studies. For individuals with fixed workweeks, recreation takes place on weekends or during pre-designated annual vacation and cannot be traded for leisure at the margin. In such cases, Bockstael et al. (1987) and Shaw (1992) suggest the opportunity cost of time no longer need be related to the wage rate. These studies suggest that both the travel cost and travel time be included as separate variables, along with their respective constraints, income and total time available for recreation.

Table 5 contains a list and definition of variables used in the TCM demand model. This study chooses its variables according to the consumer demand theory and past literature on deer hunting in California. Individuals who hunt on opening day, belong to hunting organizations, hunted in previous seasons, and had a successful deer harvest may take potentially more hunting trips because such hunters have higher preferences, experience or skill in deer hunting recreation. Since a majority of hunters in our dataset work a fixed workweek, we assume the deer hunter maximizes utility level subject to their income and time constraints (Shaw 1992). In other words, time is a constraint like income for time intensive activities like hunting. The total time budget is constructed for the TCM model according to their time information questions they answered at the end of the survey regarding vacation time and hunting schedule. In this study, the total time budget ranges from 8 to 31 days since the deer-hunting season in SJRD lasted for one month only.

Table 5 Variables included in Travel Cost Model

Variable	Definition
Dependent Variable	
NUMTRIPS	Number of primary purpose of deer hunting trips taken to the SJRD during 1999 deer hunting season.
Independent Variables	
AGE	Hunter's age
DEERKILL	Did you harvest a deer in this area during this hunting season? 1= YES, 0 = NO
HUNTOPEN	Did you hunt on opening day of the season? 1= YES, 0 = NO
HUNTORG	Are you a member of a Sportsman's organization? 1= YES, 0 = NO
PREVSEAS	Have you hunted in this area in a previous season? 1= YES, 0 = NO
PRIVLAND	Did you hunt on private land? 1= YES, 0 = NO
RTRAVMILES	Round trip travel miles from home to the hunt zone
PCINC	Hunter income
TOTIMEBUD	Total time budget during hunting season
TRAVETIME	Number of hours one-way travel time

Count Data Nature of TCM Dependent Variable

The nonnegative integer characteristic for the dependent variable, number of seasonal trips, is from a count data process. Given the count data form of the dependent variable, a preferred estimation model would be the Negative Binomial count model to estimate the demand function (Creel and Loomis, 1990). The Negative Binomial is the more generalized form of the Poisson distribution, which allows the mean of trips to be different from its variance. The count data TCM model is specified in Equation 2.2:

$$\begin{aligned} NUMTRIPS = EXP\,(&C(1) + C(2)*AGE + C(3)*DEERKILL + C(4)*HUNTOPEN + \\ &C(5)*HUNTORG + C(6)*PREVSEAS + C(7)*PRIVLAND - C(8) * \\ &RTRAVMILES + C(9)*PCINC + C(10)*TOTIMEBUD - \\ &C(11)*TRAVTIME) \end{aligned}$$

In equation 2.2, we expected the coefficient for DEERKILL (i.e. C (3)) to have a positive sign, since hunters would likely take more hunting trips if the hunting quality had been good. Also, if hunters hunt on the opening day (i.e. C (4)), private land (i.e. C (7)), and/or previous seasons (i.e. C (6)), and belong to hunting organizations (i.e. C(5)), then we expected a positive effect on the number of trips the hunter takes as these variables indicate a strong preference for the deer hunting activity. For those hunters with a higher income level (i.e. C (9)) and/or higher total time budget (i.e. C (10)) we expect more hunting trips as well due to less binding income and time constraints. However, round-trips travel distance (i.e. C (8)) and travel time (i.e. C (11)) are expect to have negative effects on the number of hunting trips because increases of these two variables increase hunter's expense.

Calculation of Consumer Surplus in TCM

The consumer surplus from deer hunting is computed from the demand curve as the difference between what people are willing to pay (e.g. the entire area under the demand curve) and what people actually pay (e.g., their travel costs). Because the count data model is equivalent to a semi-log functional form, consumer surplus from a trip is calculated as the reciprocal of the coefficient on round trip travel miles, expressed in RTRAVMILES scaled to dollars using the cost per mile (Creel and Loomis, 1990).

Hunter Survey Data

For cost effectiveness in data collection, a mail questionnaire was sent to a random sample of deer hunters with licenses for zone D19, which includes the San Jacinto Ranger District. Of 762 questionnaires mailed to deer hunters in California during the 1999 hunting season, 7 were undeliverable. A total of 356 deer hunters' responses were collected after two mailings. The response rate is, therefore, approximately 47%. Among these respondents, 69 did not hunt deer in San Jacinto Ranger District. The response rate of this study is suspected to be low because many of the hunters did not hunt in the SJRD portion of the D19 Hunt Zone, therefore, failed to return the survey.

Statistical Results

Estimation results are summarized in Table 6. There is a negative effect of travel miles, travel time, and income on number of trips taken. Income, in this study, is insignificant. The regression results of this study indicate whether a hunter successfully harvested a deer during the hunting season (i.e. DEERKILL), whether the individual hunted on opening day (i.e. HUNTOPEN), whether the hunter hunted in this area in a previous season (i.e. PREVSEAS), and total time budget (i.e. TOTIMEBUD) had a positive and significant effects on the number of hunting trips hunters take. Consistent with economic theory, hunters with longer round trip travel miles (RTRAVMILES) and greater travel time (TRAVTIME) tend to take less hunting trips.

Table 6 Estimated Negative Binomial Count Data TCM Demand Equation

Dependent Variable: NUMTRIPS				
Method: ML – Negative Binomial Count				
	Coefficient	Std. Error	Z-Stats	Prob.
Constant	1.324485	0.216326	6.122636	0.0000
AGE	0.001395	0.003787	0.368472	0.7125
DEERKILL	0.366571	0.154703	2.369516	0.0178
HUNTOPEN	0.524153	0.114843	4.564079	0.0000
HUNTORG	0.067655	0.105870	0.639036	0.5228
PREVSEAS	0.285282	0.134456	2.121759	0.0339
PRIVLAND	0.038041	0.131497	0.289295	0.7724
RTRAVMILES	-0.002230	0.000895	-2.490006	0.0128
PCINC	-1.00E-06	2.78E-06	-0.359579	0.7192
TOTIMEBUD	0.010128	0.004824	2.099444	0.0358
TRAVTIME	-0.289315	0.086776	-3.334054	0.0009
$R^2 = 0.2058$, Adjusted $R^2 = 0.1685$				
Consumer surplus = $134.53/trip				
90% confidence interval: $81.1293 ~ 393.597				
Marginal consumer surplus per deer harvested= $257.17/deer				
90% confidence interval: $154 ~ 752				

In Table 6 the consumer surplus is calculated by:
$1/\beta$ (i.e.: coefficient of distance) * $0.3/mile (i.e.: cost per mile).= 1/0.002230 *$0.3 = 448.43 * $0.3 = $134.53/trip, where the $0.3 is the thirty cents per mile the sample average cost per mile.

Finally, the 90% confidence interval in Table 6 is obtained by the following equation:
90% Confidence Interval on Consumer Surplus per Trip =
$1/(\beta_{DIST} \pm 1.64 * 0.000895)$ * $0.30/mile = $81.13 ~ $393.59 dollars per trip

Estimating the Benefits of Harvesting an Additional Deer

The average number of trips per hunter is 5.56 trips and one out of 10 deer hunters successfully harvests a deer. To calculate the incremental or marginal value of an additional deer harvest we can use the TCM demand equation to predict the extra number of trips the deer hunter would take if they knew they would harvest a deer that season. This essentially shifts the demand curve out by the amount of the coefficient on deer harvest. The equation predicts that each hunter would take 1.9116 more trips each season if they knew they would harvest a deer. Therefore, the marginal value of another deer harvested (i.e.. marginal consumer surplus) is equal to $134.53 * 1.9116 = $257.17 per deer harvested. Finally, the 90% confidence interval in Table 6 for an additional deer harvested is obtained by applying the 90% CI on the value per trip times the additional number of trips taken by the hunter: 90% Confidence Interval of the value of harvesting an additional deer =

1.9116 * $81.13 ~ 1.9116*393.59 = $155 ~ $752 dollars per deer harvested.

Benefits Of Prescribed Burning

Table 7 provides this study's bottom line--the annual deer hunting benefits of additional acres of prescribed burning. While the initial deer hunting benefit response to prescribed burning of 1,100 acres ranges from $4,112 to $8,481 depending on the model, the incremental gains for more than the current acreage of prescribed burning is quite similar across models. That is, the annual economic hunting benefits of increasing prescribed burning from its current level of 1,100 acres to 4,810 acres is $2,056, regardless of the model used. Likewise for an additional 3,700 acres of prescribed burning to 8,510 acres, the deer hunting benefits are calculated to between $1,028 to $1,285 each year, fairly similar despite the different modeling approaches.

Table 7 Annual Deer Hunting Benefits from Increased Prescribed Burning: Macro Time Series Model and GIS Micro Model Results.

RX acres burned	Additional acres Burned	Time Series Marginal increase in Deer Harvest	Annual Increase in Deer Hunting Benefits	GIS Marginal increase in Deer Harvest	Annual Increase in Deer Hunting Benefits
1	NA	NA	NA	NA	NA
1100	1110	33	$8,481	16	$4,112
4810	3700	8	$2,056	8	$2,056
8510	3700	4	$1,028	5	$1,285

Comparison to Costs

The costs of prescribed burning on the San Bernadino National Forest range from $210 to $240 per acre (Walker, 2001). This is lower total costs per acre than reported by Gonzalez-Caban and McKetta (1986), but substantially higher than the direct costs per acre for southwestern National Forests in Wood (1988). Nonetheless, if we use the $210 per acre figure, the full incremental costs of burning the first 1,100 acres would be $231,000, with each additional 3,700 acres burned costing $779,100. The deer hunting benefits represent at most about 3.4% of the total costs of the first 1,000 acres of prescribed burning. This finding can be used in two ways. First, the incremental costs of including deer objectives in the prescribed burn should not exceed $8,000, as the incremental benefits are no larger than this. Second, the other multiple use benefits such as watershed and recreation, as well as the hazard fuel reduction benefits to adjacent communities would need to make up the difference if the prescribed burning program is to pass a benefit-cost test.

Conclusion

This study evaluated the response of deer harvest and deer hunting benefits to prescribed burning in the San Jacinto Ranger District in Southern California. To estimate hunter's benefits or willingness to pay (WTP) for harvesting an additional deer the individual observation Travel Cost Method was used resulting in a mean WTP to harvest another deer of $257. With regard to the response of deer harvest to prescribed and wildfire, we compared a macro level, time-series model which treated the entire San Jacinto Ranger District as one area and a micro GIS model which disaggregated the Ranger District into the 37 hunting areas delineated by California Department of Fish and Game. The macro time-series model estimated a larger response to burning of the first 1,000 acres than the micro GIS model did, but for increases in fire beyond 1,000 acres, the two models provide nearly identical estimates.

Using the marginal willingness to pay for harvesting another deer calculated from the TCM deer harvest response to fire, yields annual economic benefits ranging from $4,112 to $8,481 for the first 1,000 acres burned. For additional acres burned, 3,700 acres, the gain is $2,056 annually, while for another additional 3,700 acres the increase ranges from $1,028 to $1,285 per year. The costs of prescribed burning on the San Bernadino National Forest range from $210 to $240 per acre. Thus the cost to burn an additional 1,100 acres is $231,000, which is an order of magnitude larger than the deer hunting benefits gained. Specifically, the deer hunting benefits of the first 1,000 acres represents about 3.4% of the total costs. Thus, the other multiple use benefits of prescribed burning such as; providing opportunities for dispersed recreation, protecting watershed as well as hazard fuel reduction to surrounding communities, would have to cover the rest. Investigating the extent of these benefits would be a logical next step in evaluating the economic efficiency of prescribed burning in the San Jacinto Ranger District.

While fire management practices have been identified as having widespread impacts on deer habitats, many other factors that affect deer habitat exist. These other factors include livestock grazing, timber harvesting, urban development, diseases and habitat loss along with annual weather patterns (CDFG, 1998). This study attempted to take into account as many factors as possible, however the amount of data and time available for modeling were a constraint.

Some future improvements in our modeling effort that may better isolate the effects of prescribed burning on deer habitat include controlling for the severity of wildfire as different fire severities will have different effects on vegetation and soils (Ryan et al, 1983). Further, including a vegetation and soils layer in the GIS model, rather than using elevation as a proxy, could improve the predictive ability of the GIS based model as well.

Subject to these caveats, this paper has demonstrated two approaches to estimate a production function relating prescribed burning to effects on deer harvest. We found positive and significant effects on deer harvest for the two GIS models and the positive impact of fire using a macro-time series model. The USDA Forest Service and California Department of Fish and Game can make use of these approaches for future cost benefit analysis of prescribed burning.

References

Bockstael, Nancy E., Strand, Ivar E. and Hanemann, W. Michael. 1987. Time and the recreation demand model. *American Journal of Agricultural Economics* 69 (May): 293- 302

California Department of Fish and Game. "An Assessment of Mule and Black-tailed Deer Habitats and Populations in California, with Special Emphasis on Public Lands Administered by the Bureau of Land Management and the United States Forest Service." *General Technical Report,* February 1998.

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

Chou Yue-Hong. "Exploring Spatial Analysis in Geographic Information Systems." Onword Press, 1997

Creel, M. D. and J. Loomis 1990. Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer hunting in California. *American Agricultural Economics Association* (May): 434-441

Creel, M. D. and Loomis, J. 1992. Modeling hunting demand in the presence of a bag limit, with tests of alternative specifications. *Journal of Environmental Economics and Management* 22: 99-113

Environmental Systems Research Institute, Inc. 1995. "Understanding GIS, the ARC/INFO Method." 380 New York Street, Redlands, CA. 92373 USA

Forsythe, Kenneth. "Using Conditional Probabilities to Evaluate Animal and Animal Product Import Requests: Part II – Background and Theory. U.S. Department of Agriculture, APHIS Technical Report. October 30, 1999.

Gibbs, P., D. Freeman, C. Harvey, R. Lawson, K. Valenzuela, D. Pumphrey. 1995. Area Analysis and Plan for Prescribed Burning. San Jacinto Ranger District, San Bernadino National Forest.

Gonzalez-Caban, A and C. McKetta. 1986. Analyzing Fuel Treatment Costs. Western Journal of Applied Forestry 1(4): 116-121.

Hellerstein, Daniel. 1992. The Treatment of Nonparticipants in Travel Cost Analysis and other Demand Models. *Water Resources Research* 28(8): 1999-2004

Hesseln, Hayley. 2000. The Economics of Prescribed Burning: A Research Review. Forest Science 46(3): 322-334.

Kahn, James R. 1995. *The economic approach to environmental and natural resources.* Second Edition. Dyrden Press, Forth Worth, TX.

Kie, John G. "Deer Habitat Use After Prescribed Burning in Northern California." Research Note PSW-369. Pacific Southwest Forest and Range Experiment Station, Forest Service, U.S. Department Agriculture. 1984 3p

Klinger, Robert C, Kutilek, Michael J, and Shellhammer, Howard S "Population Responses of Black-Tailed Deer to Prescribed Burning." Journal of Wildlife Management 53 (4) 1989: 863-870

Knetch, J.L. 1963. Outdoor Recreation Demand and Benefit. *Land Economics* 39 (November): 387-396

LDA. March 1976. A plan for California Deer. State of California, The Resource Agency, Department of Fish and Game. 5 pp. Illus.

Loft, Eric R. "Economic Contribution of Deer, Pronghorn Antelope and Sage-Grouse Hunting to Northeastern California and Implications to the Overall "Value" of Wildlife." *California Wildlife Conservation Bulletin* (11) 1998. 42 p

Loomis, John, Creel, Michael, and Cooper, Joseph. 1989. Economic Benefits of Deer in California: Hunting and Viewing Values. Institute of Ecology Report #32. College of Agricultural and Environmental Sciences, UC Davis. October

Loomis, J., D. Updike and C. Unkel 1989. Consumptive and nonconsumptive values of a game animal: the case of California deer. Transactions of the 54th North American Wildlife and Natural Resources Conference. Wildlife Management Institute, Washington DC.

Loomis, JB, Walsh, Richard G. 1997. Recreation Economic Decision: comparing benefits and costs 2nd Ed. State College, PA: Venture Publishing, Inc.

Paulek, Tom. "Santa Jacinto/Santa Rosa Mountains Deer Herd Management Plan."*California Department of Fish and Game*. Technical Report, 1989.

Ryan, Kevin C and Noste, Nonan V. "Evaluating Prescribed Fires." *Wilderness Fire Symposium*, Missoula, Mont. November 15-18, 1983.

Shaw, W. D. 1992. Searching for the opportunity cost of an individual's time. *Land Economics* 68 (February): 107-15

Shaw, W. Douglas, and Feather, Peter. 1999. Possibilities for including the opportunity of time in recreation demand systems. *Land Economics* 75, no. 4, (November): 592-602.

Updike, Douglas R, Loft, Eric R, and Hall, Frank A. "Wildfires on Big Sagebrush/Antelope Bitterbrush Range in Northern California: Implications for Deer Populations." *Symposium on Cheat grass Invasion, Shrub Die-Off and Other Aspects of Shrub Biology and Management*. Las Vegas, NV. April 5-7, 1989

Walker, Norm. 2001. Prescribed Fire Costs. Email from Norm Walker, Division Chief, San Jacinto Ranger District to Lucas Bair, Colorado State University. January 26, 2001.

Wood, D.B. Costs of Prescribed Burning in Southwestern Ponderosa Pine. Western Journal of Applied Forestry 3(4): 115-119.

Zedler, Paul H, Gautier, Clayton R, and McMaster, Gregory S. "Vegetation Change in Response to Extreme Events: The Effect of A Short Interval Between Fires In California Chaparral and Costal Shrub." *Ecology* 64 (4) 1983: 809-818.