# COST-BENEFIT ANALYSIS FOR NON-MARKET RESOURCES: A UTILITY-THEORETIC EMPIRICAL MODEL INCORPORATING DEMAND UNCERTAINTY

by

Trudy Ann Cameron
Department of Economics, UCLA

and

Jeffrey Englin
Pacific Northwest Laboratory

Discussion Paper #627
September 1991

Version: July 17, 1991
Last Access: August 23, 1991
File: C:\WP51\DOC\COSTBEN5.DOC

# COST-BENEFIT ANALYSIS FOR NON-MARKET RESOURCES: A UTILITY-THEORETIC EMPIRICAL MODEL INCORPORATING DEMAND UNCERTAINTY

by

Trudy Ann Cameron
Department of Economics, University of California, Los Angeles, CA 90024

and

Jeffrey Englin
Pacific Northwest Laboratory, Portland, OR 97232

## ABSTRACT

There exists a well-developed theoretical literature concerning the nonmarket value of public goods under uncertainty, but little research effort has been devoted to utility-theoretic empirical specifications. This paper develops an empirical model that employs the state preference model of consumer decision making. We use this model to assess willingness to pay for prevention of acid rain damage to lakes in the Northeast U.S. Our sample includes both resource users and non-users and we specifically model individual participation decisions, thereby allowing for individual risk (demand uncertainty) in the form of endogenous recreational participation probabilities. Controlling for user/non-user sample selection, we use responses to a referendum contingent valuation survey question to calibrate an indirect utility difference function. We then derive the corresponding cost-benefit quantities (individual expected consumer surplus, option price, option value, and individual willingness-to-pay loci) relevant to this application. As a by-product, our model supplies an intuitively appealing means of estimating total "non-use" (existence and bequest) resource values.

Acknowledgements:

A large number of individuals contributed valuable input into the design and implementation of the survey utilized in this study. Robert Mendelsohn deserves particular credit. The majority of this research was supported by the National Acidic Precipitation Assessment Program (NAPAP), the EPA and DOE under contract #DE-AC06-76RLO-1830. The results described here are preliminary and have not been subjected to review by any government or private agency, and our findings certainly do not reflect the policies or opinions of any such agency. We have benefitted from the helpful comments of participants at the W-133 conference in Monterey, CA, February, 1991 and from seminar feedback at UC Irvine, UCLA, the University of New Mexico, Los Alamos National Laboratory and the RAND Corporation. Laura Field provided valuable comments. Any remaining errors or omissions are solely the responsibility of the authors.

# COST-BENEFIT ANALYSIS FOR NON-MARKET RESOURCES: A UTILITY-THEORETIC EMPIRICAL MODEL INCORPORATING DEMAND UNCERTAINTY

## 1. INTRODUCTION

At least since Wiesbrod (1964), economists have acknowledged that a public good has value above and beyond its current use value. While individuals may choose not to use the good this period, they may choose to use it in the future. Many people would pay some amount (a contingent payment) for the right to use the good in the next period. In addition, some people may value the good even if they will never use it. For many private goods, rights to the use of a good in subsequent periods are sold in organized markets. This is rarely, if ever, true for public goods.

A general theoretical framework for cost-benefit analysis for a non-market good under uncertainty was first articulated clearly by Graham (1981). This framework is applicable in a wide array of circumstances, but we will concentrate on the implications of Graham's analysis for the estimation of the social value of environmental resources when individual <u>demand</u> uncertainty exists. Our example concerns individual willingness to pay (WTP) to prevent acid rain damage to 20% of all currently fishable high-altitude lakes in the Northeast United States. Intuition certainly suggests that the amount individuals would be willing to pay to prevent such damage will depend on the extent to which they participate in recreational angling. In particular, recreational anglers who fish in high altitude lakes will be impacted.

Much of the current empirical literature on the valuation of non-market environmental resources has focused exclusively upon user values, revealed ex post. Survey samples have concentrated upon current participants in activities that involve these resources, often intercepting respondents on location. For example, recreational fisheries valuation studies typically interview holders of fishing licenses or employ valuation surveys conducted in conjunction with fisheries management creel surveys at boat launch sites.

These studies use sample results to produce estimates of total social value by taking one of two approaches. In the first, the analyst assumes that the proposed change in the resource will not affect the number of users and therefore employs sample weights to inflate the individual sample value estimates to a population total value. In the second approach, a separate aggregate participation function is used. This function relates the total number of participants to current conditions. Then, for some forecasted change in the resource, the resulting change in aggregate participation is predicted, and these revised participation rates are combined with revised predictions about changes in individual values to forecast the consequences for total social value.

However, in addition to user values, economic theory has characterized at least two other classes of demand. "Option" demand reflects the fact that some potential users may not participate at present, and they are uncertain about whether or not they will participate in the future, but the resource has value to them because they derive utility from the option to participate. "Existence" demand (sometimes lumped together with "bequest" demand) acknowledges that some individuals have no intention whatsoever of using the resource themselves, but they derive utility from the fact that the resource continues to exist for others to use or for later generations to enjoy.

Most empirical studies which consider non-use demands invite survey respondents to reveal their total value of the resource and then to partition this value among use and non-use demands.[1] One problem with this approach is that survey respondents may have difficulty understanding the theoretical distinctions between these types of values. A second difficulty is the static nature of this approach. Without some way of forecasting the changes in the probability of using the resource under different circumstances, the analyst can only identify ex post measures of non-use values. Ex ante value measures are usually argued to be more relevant, since cost-benefit analyses of most policy measures and the attendant decisions must be made in advance of knowing the resolution of the uncertainty.

Any model designed to produce estimates of the use and non-use components of the social value of a non-market resource must explicitly feature uncertainty about the individual's user status.[2] Our study adopts a strategy of estimating the two structural equations underlying the standard "state preference" approach to modelling individual choice under

---

[1] Some representative papers on nonuse demand for nonmarket resources include Brookshire, Eubanks and Randall (1983), Greenley, Walsh, and Young (1981), Madariaga and McConnell (1987), McConnell (1983), Randall and Stoll (1983), Smith (1987a,b), and Walsh, Loomis and Gillman (1984). Additional aspects of the issue of non-use demands for extramarket goods are raised in surveys by Cummings, Brookshire, and Schulze (1986) and in Mitchell and Carson (1989).

[2] Uncertainty in the general context of cost-benefit analysis has received considerable attention over the last twenty years. Schmalensee (1972), Willig (1976), Chipman and Moore (1980), Hausman (1981), and McKenzie and Pearce (1982) all either touch on the subject or make it the centerpiece of their theoretical analyses.

uncertainty.[3] These structural equations include an equation describing the probability of being a user and an equation describing the indirect utility functions of users and non-users. We take as our starting point the deterministic theoretical model offered by Graham (1981).

This present research takes advantage of a unique survey of households in the Northeastern United States. Both users and non-users are included in the sample, and all respondents answer a common set of questions regarding their valuation of a set of environmental changes affecting freshwater recreational opportunities. These valuation questions employ the closed-ended (or "referendum") contingent valuation method (CVM) for eliciting values. Additional survey questions collect sociodemographic information. An independent data set provides an inventory of water-based recreational opportunities.

This paper is organized as follows. Section 2 reviews Graham's model. Section 3 discusses the design of our survey and outlines the data it provides. In Section 4, we develop an empirical version of the theoretical model. Using this framework, we then derive several welfare measures relevant for cost-benefit analysis. These include both individual and social measures: expected consumer surplus, option price, option value, willingness-to-pay locus, and the expected value of the "fair bet" point. Section 5 presents our empirical results, and a final section concludes the paper.

---

[3] Another approach, advocated by Smith (1987a,b) is to develop "planned expenditure functions." However, since the expenditure function is a reduced form, it obscures the underlying relationships. Graham's original analysis is easier to interpret and to implement in the context of utility functions.

## 2. REVIEW OF THE THEORETICAL LITERATURE

One of the most frequently cited theoretical analyses of cost-benefit procedures in the presence of uncertainty is Graham (1981). In this section, we follow Graham's approach, but focus on environmental policies.[4]

Assume that there are two goods, "dollars" (income) and a proposed change in environmental quality. For each individual, there are also two possible "states of nature": being a user, or a non-user, of the affected resource (with probabilities $P_u$ and $P_n$, respectively). Following Arrow's state preference approach, the consumer will have claims to dollars dependent upon which state occurs. Using Hirschleifer's (1965, 1966) extension of the von Neumann-Morgenstern theorem, the individual's utility function can be represented as:

$$(1) \qquad V = P_n V_n(c_n, \delta) + P_u V_u(c_u, \delta)$$

where $c_n$ and $c_u$ are claims to dollars contingent upon user or non-user status, respectively, and $\delta$ represents the presence or absence of the proposed change in environmental quality. (In our survey, some of the hypothesized environmental changes were beneficial and some were adverse. For the example in this paper, the change is adverse, so we will assess willingness to pay to avoid the change.) Let $\delta=1$ indicate that the change has occurred, and $\delta=0$ indicate that it has not happened.

Following Graham, we make the standard assumptions of nonsatiation and risk aversion (conditions upon the first and second derivatives of the utility function). We can now define "expected surplus" and "option price."

[4] Additional work on uncertainty in valuation has been done by Chavas, Bishop, and Segerson (1986) and by Chavas (1991).

Let individual contingent surplus $S_j$, j=n,u, be defined by the condition:

(2) $$V_j(Y - S_j, 1) = V_j(Y, 0); \qquad j=n,u$$

The individual's expected surplus (expected equivalent variation in our application) is:

(3) $$E[S] = P_n S_n + P_u S_u$$

The second quantity, option price, is defined by the equality:

(4) $$P_n V_n(Y - OP, 1) + P_u V_u(Y - OP, 1) = V^*$$

where

(5) $$V^* = P_n V_n(Y, 0) + P_u V_u(Y, 0)$$

Option value is then defined as the difference: $OV = OP - E[S]$.

A concept which is crucial to cost-benefit analysis under uncertainty, however, is the "willingness-to-pay locus." This is a set of ordered pairs, $(\gamma_n, \gamma_u)$ satisfying:

(6) $$P_n V_n(Y - \gamma_n, 1) + P_u V_u(Y - \gamma_u, 1) = V^*$$

where $V^*$ is as defined in (5). A consumer facing "individual risk" regarding his usage status would be happy to make any of these contingent payments ($\gamma_n$ if he turns out to be a non-user and $\gamma_u$ if he turns out to be a user) rather than suffering the adverse environmental change (or doing without the desirable environmental change).[5]

Once this locus has been identified, it is easy to see that ($S_n$, $S_u$) and (OP, OP) are two points which lie along it. The locus itself will be concave due to the assumption of risk aversion.[6] Figure 1 reproduces an adaptation of Graham's diagram. Two other interesting points on the locus are the "certainty" point and the "fair bet" points. The certainty point is defined as that point ($\gamma_n$*,$\gamma_u$*) along the willingness-to-pay locus such that:

$$V_n(Y - \gamma_n*, 1) = V_u(Y - \gamma_u*, 1) \tag{7}$$

In words, if the individual could contract for contingent payments ($\gamma_n$*, $\gamma_u$*), he would be indifferent as to whether he turned out to be a user or not. This would be a completely insured position against the individual's uncertain user status.

The fair bet point (fb) is that point ($\gamma_n'$, $\gamma_u'$) along the locus which has the largest expected value. Combinations of payments with the same expected value lie along a line with slope equal to $-P_n/P_u$. Maximizing

---

[5] The expected values of various payment combinations along the individual willingness-to-pay locus are appropriate for determining <u>individual</u> valuations under most circumstances. However, when a researcher is attempting to determine the aggregate social value of some change in environmental quality and community risk, rather than individual risk, is relevant, there will be scenarios under which the aggregate willingness-to-pay locus is an appropriate construct for analysis.

[6] Mendelsohn and Strang (1984) note that taking the total differential of equation (6) and rearranging produces the slope of the WTP locus:

$$\partial\gamma_u/\partial\gamma_n = - [P_nV_n'(Y - \gamma_n)]/[P_uV_u'(Y - \gamma_u)]$$

where $V_j'$ is the marginal utility of income for j=n,u. Diminishing marginal utility clearly produces the curvature in the locus.

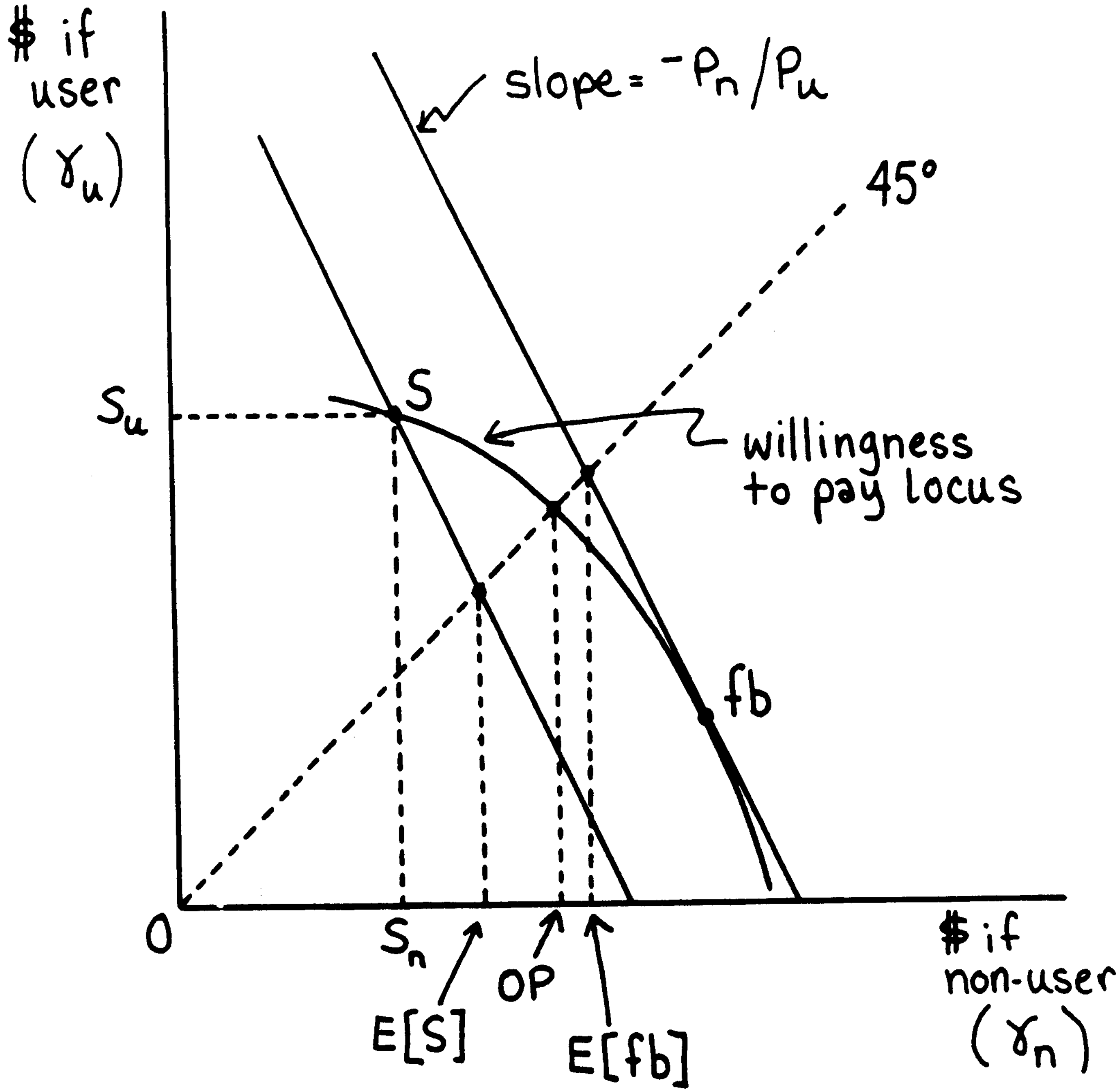


Figure 1: Cost-benefit quantities which can be read from an individual's WTP locus.

expected value along the willingness-to-pay schedule involves seeking the highest iso-expected-value line just tangent to the curve in the positive quadrant. (The tangency point gives the fair bet combination of contingent payments.)

When there exist markets in which the individual can obtain actuarily fair insurance against the uncertainty involved, the expected value of the fair bet is occasionally advocated as an appropriate measure of social value. In our context, however, such an insurance market is unlikely to exist. Adverse selection would be a serious problem, since individuals would clearly have far better information about their probability of participating next period than could any potential insurer.

There has been considerable discussion in the literature concerning which of these quantities constitute first-best measures of value for cost-benefit analyses under different circumstances. Ex ante, option price, OP, is a more relevant measure of value than ex post expected consumer surplus. Most previous studies have had to make do by identifying, at best, the location of the $(S_n, S_u)$ point and approximating the slope of the iso-expected value line $(-P_n/P_u)$ in order to locate the intersection of this line with the 45° line to determine the second-best welfare measure E[S]. The now-vast literature devoted to defining conditions under which option value will be positive or negative was motivated by the need to determine whether an empirical estimate of E[S] is an upper or a lower bound on option price (which would be the first-best measure if it were attainable). This paper demonstrates that the usual compromise second-best E[S] measures can be replaced by first-best OP measures.

## 3. AN OUTLINE OF THE AVAILABLE DATA

During the summer of 1989, the National Acidic Precipitation Assessment Program, in conjunction with the Office of Policy Planning and Evaluation of the Environmental Protection Agency, conducted a four-part survey in four states: Maine, New Hampshire, Vermont and New York. The survey design included a screening survey of the general population and three subsequent panels of freshwater recreationists identified during the screening survey. The panels included anglers, swimmers and boaters.

The overall goal of the data collection effort was to develop a sample of recreationists which could (i.) be linked back to the general population, and (ii.) be used effectively to implement a variety of non-market economic models. In this paper we focus on the screening survey of the general population, and concentrate upon the contingent valuation questions posed in that survey.

The general population screening survey utilized a stratified sample. The probability of a county being drawn was proportional to the population of the county. The five counties comprising New York City were excluded from the survey because of the low freshwater recreation participation rate. A sample of forty of the ninety-seven counties in the sample area was drawn. For a given county, a random digit dialing procedure was used to generate potential interviews. All households were eligible to participate in the screening survey. Since the proportion of participants in freshwater recreation by county was unknown in advance, the size of the screening survey could not be predicted a priori. Eventual sample size depended on the rate of recruitment into the three panels. The goal of the recruitment was to enlist twenty-eight anglers, seventeen swimmers and seventeen boaters from each county. The final screening survey completed interviews with 5,744 individuals, of which 4319 had complete data for all of the variables considered in this analysis.

The questionnaire was developed in a multi-step process. Focus groups were conducted with freshwater recreationists in two parts of the study area. The questionnaire was also pretested before being fielded. Some features of the survey design merit comment. One is that the questionnaire includes a device to avoid a female phone answering bias. The device used is to ask for the individual over eighteen with the most recent birthday. This randomizes the respondents by gender. Additionally, the screening survey contains a fairly broad set of demographic questions since it was intended to be the primary tool used to gather demographic characteristics for the full range of studies anticipated from the extended data set. These characteristics include: secondary residence, education, employment status (including retired), household size, number of children (and their ages), age, ethnicity, income, and gender. Acronyms and descriptive statistics for the subset of variables employed in this paper are presented in Table 1. A more detailed discussion of the data is provided in
Appendix I.

There were four referendum contingent valuation questions posed to respondents during the screening interviews, but we will concentrate here on just the "acid rain" question:

> **"If acid rain damaged fishing in 20% of all currently fishable high altitude lakes in the Northeast, would you be willing to pay ____ per year to prevent this?"**

Different values between $1 and $100 were randomly assigned to each

Table 1

Descriptive Statistics for Estimating Sample (n = 4319)

| ACRONYM | Description | Mean (st.dev.)<br>Full Sample | Mean (non-users)<br>Mean (users) |
|---|---|---|---|
| USER | User this season?<br>(1=yes, 0=no) | 0.1706 | 0<br>1 |
| t | Offered CV threshold<br>(dollars) | 25.33<br>(28.69) | 24.98<br>27.05 |
| WTP | WTP offered amount<br>(1=yes, 0=no) | 0.7326 | 0.7180<br>0.8033 |
| **Income (multiply imputed in estimation phase):** | | | |
| MIDINC | income bracket<br>(midpoint, $'000) | 36.33<br>(27.54) | 36.57<br>35.18 |
| **Schooling (omitted category is less than college graduate, non-trade school):** | | | |
| TRSC | trade school is<br>highest educ. attain. | 6.483e-02 | 0.06086<br>0.08412 |
| COLG | college graduate or<br>higher degree | 0.2936 | 0.2998<br>0.2632 |
| **Gender and Ethnicity:** | | | |
| FEM | female<br>(1=yes, 0=no) | 0.5291 | 0.5592<br>0.3826 |
| BLK | black<br>(1=yes, 0=no) | 0.01204 | 0.01396<br>0.002714 |
| AMIN | American Indian<br>(1=yes, 0=no) | 0.02038 | 0.01759<br>0.03392 |
| **Life-cycle variables:** | | | |
| AGE | age in years | 42.06 | 42.83 |

| | | | |
|---|---|---|---|
| | | (15.90) | 38.36 |
| RETI | retired (1=yes, 0=no) | 0.1475 | 0.1575<br>0.09905 |
| Occupational status (omitted category is full-time employment): | | | |
| PART | employed part-time (1=yes, 0=no) | 0.1236 | 0.1329<br>0.07870 |
| NOEM | not employed (1=yes, 0=no) | 0.08613 | 0.08961<br>0.06920 |
| UNEM | unemployed (1=yes, 0=no) | 0.05210 | 0.05221<br>0.05156 |
| STUD | student (1=yes, 0=no) | 0.01806 | 0.01843<br>0.01628 |
| County attributes from NORSIS data base: | | | |
| CNTY | county area (millions of acres) | 0.6632<br>(0.5457) | 0.6490<br>0.7321 |
| LILL | small lakes in county (acres,<2 acres in size) | 5.594<br>(4.503) | 5.713<br>5.016 |
| LILR | small rivers in county (acres,<66 feet wide) | 19.77<br>(22.23) | 19.21<br>22.47 |
| POP | county population (1985, millions) | 0.2026<br>(0.3145) | 0.2167<br>0.1344 |
| State of residence (NY is omitted category): | | | |
| NH | New Hampshire (1=yes, 0=no) | 0.09956 | 0.09548<br>0.1194 |
| ME | Maine (1=yes, 0=no) | 0.1466 | 0.1335<br>0.2103 |
| VT | Vermont (1=yes, 0=no) | 0.1401 | 0.1251<br>0.2130 |
| Past general fishing experience: | | | |
| PAST | Past fishing trips? (1=yes, 0=no) | 0.4987 | 0.5793<br>0.1072 |

| | | | |
|---|---|---|---|
| YRS | Years of past fishing experience | 7.738<br>(13.12) | 5.195<br>20.09 |
| Miscellaneous variables: | | | |
| SECR | secondary residence (1=yes, 0=no) | 0.1533 | 0.1502<br>0.1683 |
| URB | urban county (1=yes, 0=no) | 0.3945 | 0.4082<br>0.3284 |

individual for each referendum contingent valuation question.

A supplementary source of data was the NORSIS database. Compiled by the U.S. Forest Service, Southeast Range Station (Athens, GA), this database provides nationwide county-by-county inventories of outdoor recreational opportunities, from numbers of swimming pools to numbers of boat-launching ramps. The information is collected from a wide variety of federal and state sources. We match this data to the NAPAP screening information on a county by county basis for the forty counties sampled in our survey.

## 4. AN EMPIRICAL SPECIFICATION

### a.) The Participation Probability Sub-model

In the theoretical literature, it is quite clear that individual *subjective* probabilities determine individual expected utility. While it is certainly feasible in survey research to ask respondents directly about their participation probabilities, our survey did not attempt to elicit these probabilities.[7] For one thing, a relevant time horizon would have to be specified, and we would also have had to rely upon respondent comprehension of the notion of probability. Instead--as is often preferred in empirical research--we choose to rely upon current-period revealed preferences for participation across our sample rather than the stated

---

[7] Our survey asked respondents whether they planned any fishing trips at all between the time of the screening survey and September 30. A yes response, however, implies only that their participation probability exceeds zero. Of our 4319 respondents, 1816 indicated that they planned to fish during this subsequent period, but the subsequent waves of the survey revealed that only 977 actually went fishing at any time during the entire April 1 to September 30 sampling period. Unfortunately, these questions pertained to fishing in general, rather than fishing in any of the high-altitude lakes that are susceptible to acid rain damage.

preferences of each individual.[8]

The participation probabilities that we infer from this first submodel deserve careful attention because these probabilities play three vital roles in our model. First, they are the pathway whereby demand uncertainty enters into the model. The relative probabilities of use and non-use determine the slope of the iso-expected value lines ($-P_n/P_u$) used to convert any given pair of contingent payments to an expected value. Second, they are crucial to the process of correcting for selectivity bias in locating the true coordinates of point S ($S_n$, $S_u$). Since respondents have freely chosen to be users or non-users of the resource in question, failure to correct for the endogeneity of participation when estimating willingness-to-pay will locate point S incorrectly and therefore result in improperly positioned individual WTP loci and invalid inferences about option prices and other welfare measures. Third, these probabilities are an ingredient in constructing the individual WTP loci themselves. Each individual's locus passes through the point S, but its shape also depends on the probabilities $P_n$ and $P_u$.

In our sample, we observe the discrete outcomes of participation or nonparticipation in freshwater recreational fishing during the current

---

[8] Consistency of our empirical model with Graham's theoretical development requires that one accept fitted probabilities derived from observed participation as reasonable proxies for respondents' true but unobserved subjective participation probabilities. Certainly, if we adopt the usual interpretation that each observation in the sample represents a large number of essentially identical individuals in the population, these fitted probabilities would be useful estimates of typical subjective probabilities for that group. Since we have no alternative but to resort to stated preferences in the second (contingent valuation) phase of the modeling process, this choice anchors at least the participation submodel upon actual behavior. Even if individual subjective probability claims had been available, we would still have modelled these subjective probabilities using the same explanatory variables we employed to infer probabilities from observed participation behavior.

period. A maximum likelihood (MLE) probit algorithm is therefore appropriate for estimating fitted continuous participation probabilities for each individual. Let the individual's (latent) propensity to participate, $W_i$, depend systematically on a vector of variables, $Z_i$:

$$W_i = \theta' Z_i + \epsilon_i \tag{8}$$

where $\epsilon_i$ is distributed $N(0,\upsilon^2)$. The location and scale of measurement of $W_i$ are unobservable, so we assume that if $W_i > 0$, participation is observed and if $W_i < 0$, the individual is a nonparticipant. Standard MLE probit techniques yield estimates of the parameter vector $\theta^* = \theta/\upsilon$, and each individual's fitted continuous probability of participation, $P_u$, is given by $\Phi(\theta^{*\prime} Z_i)$; of nonparticipation, $P_n$, by $[1 - \Phi(\theta^{*\prime} Z_i)]$. $P_n/P_u$ will then be the absolute value of the slope of the iso-expected-value line for individual i.[9]

Selectivity correction is another role for the results of the participation model. If unobservable factors which make individuals

---

[9] If all of the $Z_i$ variables were free from measurement error, the standard MLE probit parameters and their covariance matrix could be considered reliable. However, the income data we have available is in bracket form, with an open-ended upper interval. Since income data figure prominently in a utility-theoretic specification, it is important to our analysis not to overstate the information in these data. Wherever the income variable appears in any of our estimations, the effective parameter point estimates and covariance matrices have been assessed for sensitivity using the technique of "multiple imputation" (Rubin (1987) and Brownstone (1991)). Grouped data estimation of the marginal log-normal distribution of income in the full sample produces a fitted distribution for log(income). Multiple imputations of the income vector are used to produce a range of point estimates and covariance matrices subject to the vagaries of our income variable. We use the means of the imputation point estimates as parameter values and the average of the imputed covariance matrices inflated by a factor of $(1 + m^{-1})$ times the sample covariance of the m different vectors of point estimates as the parameter covariance matrix.

systematically more or less likely to participate also make them likely to exhibit higher- or lower-than-expected equivalent variations for the proposed environmental changes, our parameter estimates from the second-stage contingent valuation random utility probit model outlined in the next section will be biased unless we compensate for this tendency. To effect a correction, we will borrow from the literature on selectivity bias corrections in ordinary least squares regression models, adapting the procedure to our second-stage probit specification.[10]

In most selectivity correction models, the second stage is to be estimated only for one of the two outcomes over which the selection of the sample takes place. In this study, however, we elicit contingent valuation responses from both users and non-users. Following the lucid exposition of selectivity correction given in Greene (1990, p. 744), the vector of regressors in our main contingent valuation model will be augmented by two constructed variables:

$$(9) \qquad \begin{aligned} \lambda_i^u &= \phi(\theta^{*\prime}Z_i) \,/\, \Phi(\theta^{*\prime}Z_i) && \text{if } D_i = 1 \\ &= 0 && \text{otherwise} \\ \lambda_i^n &= -\phi(\theta^{*\prime}Z_i) \,/\, [1 - \Phi(\theta^{*\prime}Z_i)] && \text{if } D_i = 0\ . \\ &= 0 && \text{otherwise} \end{aligned}$$

---

[10] Our primary results are derived using simultaneous full information maximum likelihood estimation of the participation and contingent valuation joint probit models. However, for sensitivity analysis, the estimation must be simulated many times to accommodate our deficient income variable. For our sequentially estimated results, we rely on fast single probit algorithms and correct the second stage contingent valuation probit covariance matrix.

where $D_i = 1$ for current users and $D_i = 0$ for current non-users. These two variables will be included as explanatory variables in the valuation model described in the following section. In that model, the estimated coefficients on these two selectivity correction terms will reveal information about the correlation between the effects of unobservable variables which simultaneously affect participation and resource valuation.

b.) <u>A Random Utility Model for the Indirect Utility-Difference Function</u>

We will start with the simplest utility function that seems to allow non-zero option values to be estimated--one that is essentially Cobb-Douglas in form. Let indirect utility with and without the contingent valuation scenario be:

$$(10) \quad V_j^1 = \beta_j \log(Y-t) + \delta_j \log P + \omega_j \log(A) + \alpha_j^1 X + \eta^1, \; j = \text{non-user, user};$$

$$(11) \quad V_j^0 = \beta_j \log(Y) + \delta_j \log P + \omega_j \log(.8A) + \alpha_j^0 X + \eta^0, \quad j = n, u.$$

where $V^1$ implies the respondent's indirect utility <u>with</u> the hypothesized change in the resource (prevention of acid rain damage) and the loss of income (the payment, t) proposed in the contingent valuation question. $V^0$ implies the situation with <u>no</u> mitigation of acid rain and no payment. Y is income, P is the price of a day of access to the specific environmental resource under consideration, A is the attribute of the resource which is subjected to a hypothetical change (acid rain damage to 20% of all currently fishable high-altitude lakes, leaving the fishable proportion at 0.8 of the current number), and X is a vector of respondent sociodemographic characteristics, other resource attributes, and other prices. The assumed i.i.d. normal error terms are $\eta^1$ and $\eta^0$.

The respondent's answer to the contingent valuation question will depend on the relative magnitudes of these two utilities. They will be willing to pay the proposed amount if the utility difference (10)-(11) is positive:

$$
\begin{aligned}
(12)\quad V_j^1 - V_j^0 = \beta_j \log\left(\frac{Y - t}{Y}\right) - \omega_j \log(.8) + (\alpha_j^1 - \alpha_j^0)'X \\
+ \rho_j \lambda^j + (\eta^1 - \eta^0) > 0, \\
j = n,\ u.
\end{aligned}
$$

We will rename the $(\alpha^1{}_j - \alpha^0{}_j)$ coefficients as simply $\alpha_j$ (since we will be unable to identify their separate values). The term involving the $\omega_j$ parameter and the log of (.8) will be absorbed into the constant term among the $\alpha$ coefficients.[11] The $\lambda$ terms are the selectivity correction variables defined in (9) in the previous section. If we assume that the error terms in the participation probit model and the upcoming contingent valuation probit model are distributed bivariate normal with parameters $[0,0,1,1,\rho]$, a simple analogy to the ordinary least squares case would imply that the coefficient on each $\lambda$ variable should be a separate estimate of the correlation parameter $\rho$. This approach will produce two separate estimates of $\rho$: $\rho_u$ and $\rho_n$, which can be examined for comparability.[12]

[11] If our questionnaire had varied the hypothesized level of damage to high-altitude lakes across respondents, we could readily have determined a schedule of WTP a function of damage levels prevented. Since only the 20% damage level was addressed, we produce only a point on this curve.

[12] In the sequentially estimated model, the estimates of $\rho$ can be constrained to be identical by summing the two separate $\lambda$ variables to create a single selectivity correction term. The value of $\rho$ is estimated

For current <u>users</u>, we must use our calibrated model to simulate their likely non-use value $S_n$, whereas for current <u>non-users</u>, we must use the model to simulate their likely use value, $S_u$. Only then can we locate the point $S = (S_n, S_u)$ for each individual and construct their WTP locus. If selectivity <u>is</u> occurring, and we ignore it, we will be plotting S in the wrong place.[13]

The specification in (12) follows that adopted by Hanemann (1984) in that the stochastic structure assumes that the <u>indirect utility difference</u> bears an additive $N(0,\sigma^2)$ error term, $\eta = \eta^1 - \eta^0$. We will subsume the non-user and user indirect utility functions within one model by allowing each coefficient in the utility-difference function to differ systematically across the two groups. Thus, we modify our notation so that $\beta_j = \beta_n + \beta_\delta D_i$, where j = n, u, and $D_i$ equals 0 for non-users and 1 for users. The coefficient $\beta_\delta$ therefore denotes the coefficient differential between the two groups. Likewise, we will have $\alpha_j = \alpha_n + \alpha_\delta D_i$. In its simplified form, for each individual i, the stochastic model is:

$$(13) \quad (V^1 - V^0)_i = (\beta_n + \beta_\delta D_i) \log((Y_i - t_i)/Y_i) + (\alpha_n + \alpha_\delta D_i)'X_i + \rho_n \lambda_i^n + \rho_u \lambda_i^u + \eta_i$$

Note that logP (the own-price term) disappears from the utility

---

directly by maximum likelihood in our bivariate probit model.

13 For example, if $\rho$ is positive so that unobservable factors make users systematically more likely to have high surplus and nonusers more likely to have lower surplus, our S points without a selectivity correction will tend to be placed too far to the northwest in Graham's diagram. If $\rho$ is negative, they will be too far to the southeast.

difference equation because price is presumed not to be affected by the hypothesized change. Likewise, prices of all other goods are also presumed to remain constant and these will not appear in the utility-difference function either.

Conventional packaged maximum likelihood probit models can be used to estimate this contingent valuation utility-difference model. The simulated (scaled) utility difference $(V^1 - V^0)_i$ in the absence of selectivity would be calculated by imposing $\rho_n = \rho_u = 0$, which is equivalent to ignoring the $\rho$ and $\lambda$ terms in any subsequent calculations involving $(V^1 - V^0)_i$.

The separate "true" values of $\beta_n$, $\beta_\delta$, $\alpha_n$, and $\alpha_\delta$ are not identified, since in the likelihood function, they always appear in ratio to the unknown error variance, $\sigma$. We must be satisfied instead with estimating $\beta_n^* = \beta_n/\sigma$, $\beta_\delta^* = \beta_\delta/\sigma$, $\alpha_n^* = \alpha_n/\sigma$ and $\alpha_\delta^* = \alpha_\delta/\sigma$. Fortunately, this is not a limitation because the important quantities required for our welfare calculations involve only ratios of the $\beta_n$, $\beta_\delta$, $\alpha_n$ and $\alpha_\delta$ parameters. The implicit $\sigma$ terms in the denominators of the maximum likelihood probit parameters will cancel.

c.) <u>Calculating Expectations of Individuals' Cost-Benefit Quantities</u>

An important complication in this specification is the required transformations of the estimated parameters and the error term. All of these are random variables, and our eventual estimates of the different welfare quantities are functions of these random parameters and the error term. We will assume that the maximum likelihood estimates of $\beta_n^*$, $\beta_\delta^*$, $\alpha_n^*$ and $\alpha_\delta^*$ are distributed approximately multivariate normal, with an asymptotic variance-covariance matrix produced from the expected Hessian evaluated at the optimal parameter values. Also by assumption, the standardized probit error term, $\eta^* = \eta/\sigma$, is N(0,1) and independent both across observations and from the estimated coefficients.[14]

The value of t which makes equation (13) exactly zero is the surplus (equivalent variation), $S_j$, associated with the proposed mitigation of acid rain damages. If we carry along the error term, $\eta^*_i$ as we solve the estimated version of equation (13) for the value of $t_i = S_{ji}$ which makes the utility difference exactly zero for individual i, we pass through the following intermediate steps. Simplify the notation temporarily by letting $\beta^{*\prime}D_i = \beta_n^* + \beta_\delta^* D_i$. Then

$$(14) \qquad f_j = \log[(Y_i - S_{ji})/Y_i] = (\alpha^{*\prime}X_i)/(\beta^{*\prime}D_i) + \eta^*_i/(\beta^{*\prime}D_i), \qquad j = n, u.$$

Any linear combination of normal random variables is also normal, so $\alpha^{*\prime}X_i$ and $\beta^{*\prime}D_i$ and $\eta^*_i$ will all be approximately normally distributed. However, to calculate $S_j$, we must exponentiate $f_j$. Each individual's fitted expected value of $S_{ji}$ will be given by:

$$(15) \qquad E\,(S_{ji}) = Y - Y\,E\{\,\exp[(\alpha^{*\prime}X_i)/(\beta^{*\prime}D_i)]\,\exp[\eta^*_i\,/(\beta^{*\prime}D_i)]\,\},$$

$$= Y - Y\,E\{\,\exp(f_j)\,\}.$$

No analytical solution for this complicated expectation is available.

Instead, we use the simulation method suggested in another context by

---

[14] As noted previously, it is necessary in the sequentially estimated model to correct the second-stage parameter estimates and parameter covariance matrix to reflect the grouped income data and the estimated nature of the selectivity correction terms.

Krinsky and Robb (1986). Their technique suggests that we adopt a multivariate normal distribution for the estimated probit parameters, with variance-covariance matrix as produced by the maximum likelihood probit algorithm (and modified to account for the grouped income data and the estimated nature of the $\lambda$ terms for selectivity correction.) Combined with this, we will assume an independent unit normal distribution for $\eta^*$. A large number of "random draws" for ($\beta_n^*$, $\beta_\delta^*$, $\alpha_n^*$, $\alpha_\delta^*$, $\eta^*$) will be produced and, in conjunction with the observed data, simulated distributions for $S_{ji}$ (and subsequently, for all of the other cost-benefit quantities to be discussed) will be calculated for each individual. The means of these simulated values will be taken as the expected value of $S_j$ for each individual. Note that these two expected values are inputs into the calculation of E[S] for the individual, which is a different entity since the expectation is taken not across the randomness of the parameters and errors, but across the two uncertain outcomes for that individual: use or non-use.

The fitted (expected values of) $S_n$ and $S_u$ for each individual are not the final objective of this study. For example, in the computation of option value, we must carry the stochastic properties of the estimates through our derivations of both E[S] and OP, calculate their individual expectations for each respondent, and then compute the difference for each person.

The individual consumer's expected equivalent variation from the proposed acid rain mitigation, E[S], also depends on the respondent's probability of using of the resource. We will employ fitted probabilities of participation and non-participation from our first-stage participation probit model. These fitted probabilities are also random variables because they are constructed as nonlinear functions of the estimated probit $\theta^*$ parameters: $P_u = \Phi(\theta^{*\prime}Z_i)$ and $P_n = [1 - \Phi(\theta^{*\prime}Z_i)]$. In addition to the Krinsky-Robb simulations for the $S_j$ values, we must employ the same types of simulations for $P_j$. Expected surplus across user and non-user states of the world, substituting the formulas derived above for $S_u$ and $S_n$ and simplifying, will be:

$$(16) \qquad E[S] = Y \{ 1 - ( P_u \exp[f_u] + P_n \exp[f_n] ) \}$$

In words, E[S] is a specific fraction of income determined by a probability weighted average of the exponentiated $f_j$ functions.

To determine the option price associated with the proposed change in the resource attribute, we need to work in the environment of utility differences, as in equation (12). Option price, OP, solves:

$$(17) \qquad P_u V_u^0 + P_n V_n^0 = P_u V_u^1 + P_n V_n^1$$

or, equivalently:

$$(18) \qquad P_u ( V_u^1 - V_u^0 ) + P_n ( V_n^1 - V_n^0 ) = 0.$$

From equations (13) and (14), we have expressions which can be substituted for the utility difference terms in parentheses in (18).

$$(19) \qquad (V_j^1 - V_j^0 ) = \beta_j \log[(Y-t)/Y] - \beta_j f_j, \quad j = n,u.$$

where the $\beta_j$ parameters are random and the $f_j$ terms are random variables involving the estimated random utility CV probit parameters and the standard normal error term from that model as in equation (14). We can now rewrite equation (18) as:

$$(20) \quad P_n\{ \beta_n \log[(Y-S_n)/Y] - \beta_n f_n\} + P_u\{ \beta_u \log[(Y-S_u)/Y] - \beta_u f_u\} = 0.$$

We then solve for the value of option price $OP = S_n = S_u$ that makes this equality true:

$$(21) \quad OP = Y \{ 1 - \exp[(P_n\beta_n + P_u\beta_u)^{-1} \ \{P_n\beta_n f_n + P_u\beta_u f_u\} ] \}$$

Whereas E[S] was a fraction of income determined by a probability-weighted average of the exponentiated $f_j$ functions, OP is a fraction of income determined by the exponentiated value of a probability- and $\beta$ coefficient-weighted average of the same $f_j$ functions. (Note the difference in the order of the exponentiation and the averaging.) Like the other cost-benefit quantities, the expected value of OP for each respondent will be simulated from a set of Krinsky-Robb draws from the joint distribution of ($\beta_n$*, $\beta_\delta$*, $\alpha_n$*, $\alpha_\delta$*, $\eta$*) for the $f_j$ and $\beta_j$ terms and an analogous set of draws from the joint distribution of the $\theta$* parameters for $P_j$, j = n, u.

Even if $\beta_u = \beta_n$ in this model, E[S] and OP will differ systematically. Making this assumption for the moment, recall that option value is defined as OV = OP - E[S]. In this simple case, we will have:

$$(22) \quad OV = \hat{Y} \{ [ P_n \exp(f_n) + P_u \exp(f_u) ] - \exp [P_n f_n + P_u f_u] \}$$

The sign of option value will depend upon the relative magnitudes of the two terms inside the braces. Of course, $P_u + P_n = 1$, so in this case, we are comparing a weighted average of two exponentiated quantities with the exponentiated value of their weighted average, with the same weights used. Since the function exp(z) is convex when viewed from below, the first term will always exceed the second term, regardless of the values of the two $f_j$ functions. Thus if $\beta_u = \beta_n$, option value will be positive.

Only if the weights used in the expression for OP (i.e., in equation (21)) differ from the simple probability weights in the expression for E[S] is it possible for option value to be negative for some or all consumers. Suppose that $\beta_u = \beta_n + \Delta$. We can substitute this expression for $\beta_u$ wherever it appears in equations (16) and (21), set E[S] = OP and solve for the value of $\Delta$ that makes the equality true:

$$(23) \qquad \Delta = -\beta_n \{ 1 + P_n[\log((Y-E[S])/Y) - f_n]/P_u[\log((Y-E[S])/Y) - f_u] \}.$$

The threshold $\Delta$ will differ across individuals due to differences in $P_j$, $f_j$ and Y. If $\Delta_i < \beta_u - \beta_n$, individual i's option value will be positive; if not, it will be negative.

The issue of option value was a preoccupation in the literature for some time, so we will provide empirical estimates of option value in our discussion of the results. However, option value is not a distinct component of value, and OV estimates have limited relevance. OV is merely the difference between an ex ante value measure (option price) and an ex post measure (expected consumer surplus). The only time option value may be useful is for assessing the error introduced when using an available E[S] estimate as a proxy for the desired OP measure (see Smith, 1987a, p. 289).

Having shown how the various welfare measures are derived, we are now in a position to examine one of the limitations of our current specification. Our model makes explicit the endogeneity of participation status in the contingent valuation indirect utility-difference equation. We also allow for unobserved heterogeneity to affect both the utility difference and the participation decision through their correlated error terms. The next generation of specifications will explore more general functional forms for indirect utility and will seek to make resource values an explicit endogenous variable in the participation equation as well.

Making values a determinant of participation appears to be non-trivial in the present case because the contingent valuation questions in our survey use a referendum format. The discrete contingent valuation response explained by the random utility model cannot be used analogously in the participation equation due to its dependence on the offered threshold. For our indirect utility function, each individual's option price, for example, is not an observed or directly estimated variable but an expected value inferred through simulations with no simple analytical formula that could be substituted into the participation equation for general full information maximum likelihood (FIML) estimation. Generalizing our model to one where valuation and participation are jointly determined will be an important undertaking in subsequent research. Our present formulation is a recursive model with a non-diagonal error covariance matrix rather than a fully simultaneous, dependent-error specification.[15]

d.) <u>Empirical Formulas for Individual WTP Loci and Useful Quantities</u>

For our specification for indirect utility, each individual's WTP locus will be given by $(\gamma_n, \gamma_u)$ pairs which satisfy the following equality:

$$(24) \quad \gamma_u = Y - Y \exp\left( (P_n\beta_n f_n + P_u\beta_u f_u)/(P_u\beta_u) - (P_n\beta_n/P_u\beta_u) \log[(Y-\gamma_n)/Y] \right)$$

In most theoretical analyses, the individual WTP locus is depicted with pronounced concavity (following directly from the assumption of risk aversion on the part of consumers). Nonsatiation requires that $\partial V_j/\partial Y > 0$ for $j = n, u$. Risk aversion requires $\partial^2 V_j/\partial Y^2 < 0$ for $j = n, u$. For our simple utility functions, the first derivative with respect to income is just $\beta_j/Y$; the second derivative is just $-\beta_j/Y^2$. Thus the conditions will be met globally as long as $\beta_j > 0$ for $j = n, u$. Furthermore, the Arrow-Pratt measure of absolute risk aversion in our case reduces to simply $1/Y$ (a very tiny number) which is independent of the parameter estimate for $\beta_j$. While the correct sign on the $\beta$ coefficients will guarantee concavity of the WTP locus, it is important to appreciate that the curvature of the locus in

[15] A naive simultaneous formulation of our two equations incorporates both the contingent valuation discrete response <u>and</u> the offered threshold value as explanatory variables in the participation probit model. Estimating the participation model by itself, the threshold variable is strongly significant and the discrete response to the valuation question is significant at the 10% level. However, when the two equations are estimated jointly using the bivariate probit algorithm, the estimated correlation between the two equation errors persists in moving outside the admissible range and the algorithm does not converge. This outcome is common in simultaneous bivariate probit models.

our empirical examples will be very slight.[16]

Consider the algebraic formula for the slope of the WTP locus:

$$(25)\qquad |\partial\gamma_u/\partial\gamma_n| = [Y/(Y - \gamma_n)]\ (P_n\beta_n/P_u\beta_u)\ \cdot$$
$$\exp(\ (P_n\beta_n f_n + P_u\beta_u f_u)/(P_u\beta_u)\ -\ (P_n\beta_n/P_u\beta_u)\ \log[(Y - \gamma_n)/Y]$$

This derivative is clearly increasing in $\gamma_n$, but in our application, the contingent payments ($\gamma_n$, $\gamma_u$) are typically very, very small relative to income Y. Thus $[(Y - \gamma_n)/Y] \approx 1$ and the slope will be almost a constant (relative to $\gamma_n$) given by:

$$(26)\qquad |\partial\gamma_u/\partial\gamma_n| = (P_n\beta_n/P_u\beta_u)\ \exp(\ (P_n\beta_n f_n + P_u\beta_u f_u)/(P_u\beta_u)\ )$$

Anticipating future efforts to use models of this genre to estimate resource values, it should be emphasized that it is crucial to have data on participation in an activity which is as closely matched as possible to the resource change described in the contingent valuation question being analyzed. The shape of the individual WTP loci depend crucially upon the odds of being a non-participant ($P_n/P_u$); if these odds are misrepresented by an inappropriate match between the definition of participation and the CV scenario to be examined (e.g. participation defined too broadly), the loci

---

16 Graham (1984) rebuts Mendelsohn and Strang (1984) by asserting that their claim that "...projects which entail individual risks and 'small' changes in the marginal utility of income across states give rise to linear WTP loci..." is incorrect. Our empirical findings suggest that linear WTP loci are a rather good approximation, at least under our specification for indirect utility. However, this approximately constant slope is not equal to $-P_n/P_u$.

will be incorrect (in this case, too flat).

We have already noted that the expected value of the fair-bet point is probably not appropriate in this particular context as a measure of welfare because it is unlikely that actuarily fair insurance against the use/non-use contingency would ever be offered to individuals. However, if this insurance market did exist, we could readily calculate this welfare measure. Once each individual's WTP locus is identified, their fair bet point is found by setting the absolute value of the slope of the WTP locus equal to the individual's current relative probability of non-use, $P_n/P_u$. This yields the fair bet coordinates $(\gamma_n', \gamma_u')$, where

$$\gamma_n' = Y - Y \exp( (P_n\beta_n f_n + P_u\beta_u f_u)/(P_n\beta_n + P_u\beta_u) + (P_u\beta_u)/(P_n\beta_n + P_u\beta_u) \log(\beta_n/\beta_u) ) \tag{27}$$

and $\gamma_u'$ satisfies equation (24) for the WTP locus. These formulas apply providing that the tangency of the iso-expected value line with the WTP curve is interior. Corner solutions are frequent, however. For $j = n, u$, if $\gamma_j' < 0$ according to the above formulas, we substitute $\gamma_j' = 0$ and

$$\gamma_j' = Y - Y \exp ( (P_n\beta_n f_n + P_u\beta_u f_u)/(P_j\beta_j) ) \tag{28}$$

The expected value of the fair bet point for each consumer employs the individual fitted participation probabilities for that person:

$$E[fb] = P_n \gamma_n' + P_u \gamma_u' \tag{29}$$

The "certainty point" along the WTP locus would be an interesting curiosity, since this combination of contingent payments allows the respondent to be fully insured against risk involving his user/non-user status. However, this particular welfare quantity is not available given the specification used in this study. Given that the parameters of the indirect utility function in this application must be estimated from the indirect utility-difference equations, we cannot recover all the information necessary to identify these coordinates. Referring back to section 2 and using the indirect utility functions outlined at the beginning of this section, the certainty point is the pair of contingent payments $(\gamma_n^*, \gamma_u^*)$ along the WTP locus that satisfies:

$$(30) \qquad 0 = V_u^1 - V_n^1 = \beta_u \log[Y - \gamma_u^*] - \beta_n \log[Y - \gamma_n^*] - (\omega_u - \omega_n) \log(0.8A) + (\alpha_u^1 - \alpha_n^0)\, X$$

While point estimates of the $\beta$ coefficients are available, as are data for Y and the X variables, it is not possible to solve this equality for $(\gamma_n^*, \gamma_u^*)$.[17]

[17] For this functional form, no distinct estimates of the usage-status-specific $\alpha$ parameters are available (we estimate only $(\alpha^1_u - \alpha^0_u)$ and $(\alpha^1_n - \alpha^0_n)$. Second, the level of the environmental amenity, A, enters in log form so that its magnitude conveniently drops out of the utility difference function that is estimated. This is fundamentally necessary because no data on the absolute level are available. The constant $-\omega_j \log(0.8)$, $j = n, u$, is absorbed by the constant term in each utility difference function in the estimation process, so neither term in the second line of the formula can be recovered.

## 5. EMPIRICAL RESULTS

### a.) The participant/nonparticipant probit submodel

Many hundreds of individual lakes in the Northeast U.S. are susceptible to acid rain damage, but the state of the science does not allow researchers to pinpoint precisely which lakes would be among the first 20% to be affected. Hence our contingent valuation question did not list specific lakes which would allow us to select only users of those particular lakes as "participants" for purposes of the first probit model in our analysis. However, acid rain is far more likely to adversely affect higher altitude, shallow bodies of water. In order to match users as closely as possible with the lakes addressed in the CV question, our definition of participants is limited to those individuals who actually went fishing at least once during the sample period and who also listed some type of trout as one of their targetted species. Anglers who target bass exclusively are not members of the intended user group, for example.

Table 2 describes parameter estimates and asymptotic t-ratios for our probit model of participation in fresh-water recreational fishing in high-altitude lakes in the Northeast U.S. Two sets of point estimates and standard errors are provided. The first set is for the participation portion of a bivariate probit model where the participation outcome and the contingent valuation response are estimated jointly. The second set gives point estimates and modified standard errors that result from 50 multiple imputations of the grouped data income variable. Most coefficients bear the anticipated sign. Three categories of variables are employed in calibrating the participation submodel. First, we use characteristics of the individual respondent. The model is specified as fully quadratic in AGE and INC

Table 2

Fishing Participation Probit Model

Estimated both as (i.) one equation in a joint probit specification, and (ii.) derived from ordinary probits using 50 multiple imputations of the income variable)

Dependent variable: USER

| Value | Label | Count | Percent |
|---|---|---|---|
| 0 | no | 3582 | 82.94 |
| 1 | yes | 737 | 17.06 |

| Independent Variable | Bivariate Probit Model (joint with CVM response) | | Single Probits on 50 imputations of income | |
|---|---|---|---|---|
| | Est. Coef. | t-Stat. | Est. Coef. | t-Stat. |
| INTERCEPT | -0.5242 | -1.663 | -0.5309 | -1.815 |
| AGE | -0.02803 | -1.966 | -0.02726 | -2.115 |
| AGE2 | 1.346e-04 | 0.827 | 1.265e-04 | 0.878 |
| INC | 0.002899 | 0.651 | 0.002760 | 0.621 |
| INC2 | -8.031e-06 | -0.745 | -8.147e-06 | -0.708 |
| AGE*INC | -5.200e-05 | -0.585 | -5.191e-05 | -0.587 |
| SECR | -0.09645 | -1.241 | -0.09473 | -1.210 |
| FEM | -0.1461 | -2.339 | -0.1474 | -2.425 |
| RETI | 0.1174 | 0.770 | 0.1169 | 0.827 |
| BLK | -0.5797 | -1.415 | -0.5922 | -1.566 |
| AMIN | -0.02276 | -0.139 | -0.01498 | -0.088 |
| PART | -0.1891 | -1.840 | -0.1892 | -1.924 |
| NOEM | 0.1365 | 1.239 | 0.1352 | 1.264 |
| UNEM | 0.04550 | 0.332 | 0.04386 | 0.342 |
| STUD | -0.07536 | -0.366 | -0.07028 | -0.339 |
| URB | 0.06184 | 0.797 | 0.06373 | 0.813 |
| CNTY | -0.1161 | -1.359 | -0.1150 | -1.372 |
| LILL | 0.01218 | 1.698 | 0.01247 | 1.707 |
| LILR | 0.0010 | 0.572 | 0.001078 | 0.602 |
| POP | -0.2543 | -1.884 | -0.2648 | -1.921 |
| VT | 0.5877 | 6.677 | 0.5893 | 6.961 |
| NH | 0.3297 | 3.302 | 0.3278 | 3.436 |
| ME | 0.5192 | 4.750 | 0.5133 | 5.001 |
| PAST | -0.9230 | -13.132 | -0.9297 | -13.66 |
| YRS | 0.08637 | 15.493 | 0.08588 | 15.79 |
| YRS2 | -0.001112 | -10.076 | -0.001106 | -9.429 |

(income), but only the linear term in AGE is individually statistically significant at the usual 5% level.

There are also several dummy variables: ownership of a secondary residence (SECR) decreases participation and retired status (RETI) increases participation, although neither effect is significant. Participation is insignificantly lower for blacks and American Indians, but it is significantly lower for females (FEM).

Part-time employment, relative to full-time employment, results in an almost significant decrease in participation. Being unemployed, or out of the labor force (as opposed to retired), both increase participation, although not significantly. Student status produced a negligible decrease in participation.

A second category of variables attempts to capture the extent of water-based recreational opportunities available in the respondent's home county. Participation is insignificantly lower for residents of urban counties (URB). CNTY is the total county area, in acres. LILL is acres of water in bodies less than two acres, LILR is acres of water in rivers or streams less than 66 feet wide, and POP is the county population estimate for 1985. Only the acreage of small lakes (LILL) comes close to having an individually significant effect among the water area variables. Being from a more densely populated county, controlling for county area, has a significantly negative effect upon participation.

Relative to New York State (the omitted category), participation by residents of the states of NH, ME and VT all exhibit statistically significantly higher probabilities of participation.

The last set of variables attempts to control for past fishing participation of any type. Whether or not an individual has fished in the past, and how intensively, is an important indicator of the probability that they will participate in the current year. The past participation dummy alone bears the opposite sign from that anticipated, but one must keep in mind that the effect of past participation is captured simultaneously by all three of the variables (PAST, YRS, and YRS2). If PAST is not zero, all three of these terms are activated. All three variables are highly significant.

These past participation variables describe "fishing in general," and could be expected by themselves to be excellent predictors of current participation in general fishing. However, we include other sociodemographic factors because we are modeling "fishing in high-altitude lakes" rather than general fishing. The inclusion of past participation (a lagged, partially endogenous variable) renders our specification dynamic and complicates our forecasting exercises. However, it has the potential to dramatically increase the accuracy of our probability estimates.

b.) The Referendum Contingent Valuation Probit Model

Table 3 gives the contingent valuation random utility probit parameter estimates and asymptotic standard errors for the jointly estimated bivariate probit participation and valuation models. It also gives analogous results for the sequentially estimated model that reflects the deficiency in the income variable, with all estimates corrected for both the grouped income data and the estimated selectivity correction terms. These estimates of the indirect utility difference function warrant attention primarily because they show how the user and non-user indirect utility functions differ. The basic theory behind the model allows for systematically different utility

Table 3

Indirect Utility-Difference Function Parameter Estimates

Referendum CV Question: WTP to Prevent Acid Rain Damage to 20% of all Currently Fishable High-Altitude Lakes in the N.E.

Dependent variable: WTP offered amount?

| Value | Label | Count | Percent |
|---|---|---|---|
| 0 | no | 1155 | 26.74 |
| 1 | yes | 3164 | 73.26 |

| | From Bivariate Probit (with participation) | | From 50 imputations of Y and the selectivity terms | |
|---|---|---|---|---|
| | $\beta_n^*, \alpha_n^{*a}$ | $\beta_\delta^*, \alpha_\delta^{*b}$ | $\beta_n^*, \alpha_n^{*a}$ | $\beta_\delta^*, \alpha_\delta^{*b}$ |
| Independent Variable | Coef. (t-ratio) | Coef. (t-ratio) | Coef. (t-ratio) | Coef. (t-ratio) |
| log[(Y-t)/Y] | 140.89<br>(13.92) | -23.86<br>(-0.763) | 142.78<br>(9.218) | -21.88<br>(-0.586) |
| INTERCEPT | 0.9889<br>(10.77) | 0.1645<br>(0.675) | 1.030<br>(10.52) | 4.139e-02<br>(0.155) |
| SECR | 0.1534<br>(2.288) | -0.1673<br>(-1.011) | 0.1577<br>(2.372) | -0.1673<br>(-1.042) |
| TRSC | 0.1786<br>(1.733) | -0.2561<br>(-1.141) | 0.1827<br>(1.829) | -0.2560<br>(-1.161) |
| COLG | 0.1947<br>(3.867) | -0.1054<br>(-0.778) | 0.1940<br>(3.855) | -0.1078<br>(-0.8111) |
| AGE | -9.237e-03<br>(-4.741) | 5.336e-03<br>(0.926) | -9.455e-03<br>(-4.742) | 5.273e-03<br>(0.9553) |
| RETI | -0.1132<br>(-1.328) | -0.1334<br>(-0.533) | -0.1161<br>(-1.379) | -0.1340<br>(-0.5518) |
| UNEM | -0.2103<br>(-2.015) | 0.5950<br>(1.952) | -0.2156<br>(-2.092) | 0.5984<br>(2.010) |
| NOEM | -0.1242<br>(-1.480) | 0.3259<br>(1.239) | -0.1258<br>(-1.519) | 0.3225<br>(1.264) |

| | | | | |
|---|---|---|---|---|
| BLK | -0.3387 (-1.807) | -[c] | -0.3493 (-1.893) | -[c] |
| FEM | 0.1555 (3.271) | -2.462e-02 (-0.190) | 0.1498 (3.147) | -3.549e-02 (-0.271) |
| $\rho$ | -0.07858 (-1.069) | - | - | - |
| $\lambda^n$ | - | - | 5.310e-03 (0.1504) | - |
| $\lambda^u$ | - | - | 3.815e-02 (0.2395) | - |

[a] Coefficients for the utility-difference function for non-users.

[b] Coefficient differentials for the utility difference function for users (i.e., add this coefficient to the one in the non-user column to yield the coefficient for users)

[c] The very small number of black respondents in this category made it inappropriate to estimate a coefficient differential for this group

functions for users and nonusers, and our model retains this feature.

In Table 3, the first column of estimated coefficients in the pair of columns for each estimation method pertains to the non-user indirect utility function (i.e., $\beta_n$* is the coefficient on log[(Y-t)/Y] and the $\alpha_n$* are the coefficients on the remaining variables). The second column of point estimates in each pair contains the differentials $\beta_\delta$* and $\alpha_\delta$* to be added to the first column to yield the estimated indirect utility parameters for users ($\beta_u$* and $\alpha_u$*). The coefficient $\rho$ is the estimated error correlation for the bivariate probit model. The estimated coefficients on $\lambda^u$ and $\lambda^n$ are $\rho_u$ and $\rho_n$, respectively, for the sequential method.

The coefficient $\beta_n$*, the marginal utility of log income for non-users, is very strongly statistically significant at about 141 to 143. The corresponding marginal utility for users bears a point value smaller by 22 to 24, but it is not significantly different from the value for non-users.

Classes of explanatory variables in the "X" category (sociodemographic shifters of the indirect utility function) include ownership of a secondary residence, educational attainment (omitted category is less-than-college graduate), age, and dummy variables for race (omitted category is white), gender (omitted category is male), and employment status (omitted category is full-time employment). For non-users, the utility difference appears to increase with education level (trade school not significant, but college significant). AGE is individually significant and the point estimate implies that the utility difference is declining with age over the ages represented in our sample. Retired status decreases the utility difference, although not significantly. The utility difference is higher if a secondary residence is maintained. Utility is statistically significantly lower for blacks (albeit only at the 10% level), but higher for females. The latter result is somewhat surprising because Table 2 shows that women are less likely to participate in fishing. While the state dummies (ME, VT, and NH) had a dramatic effect on participation, they had uniformly small and insignificant effects in the valuation portion of the model.

The second column in each pair in Table 3 shows that few of the user utility-difference parameters are individually significantly different from the corresponding non-user parameter at the 5% level. The major exception is UNEM, unemployed status. For non-users unemployment decreases the utility-difference significantly, but for users unemployment has a statistically significant positive effect. The results for the not employed dummy variable are similar, although not statistically significant.

In this particular example, the selectivity problem does not appear to be as important as it is in the analysis of some of the other three contingent valuation questions posed on our survey.[18] In the bivariate probit version of the model, the single point estimate of the error correlation $\rho$ is very small, negative and insignificant. In the sequentially estimated version, neither constrained nor unconstrained estimates of the $\rho$ parameter on the $\lambda$ terms were statistically significantly different from zero.[19] Since our example is intended as a prototype for future applications, it is important to emphasize that selectivity bias

---

[18] The researcher cannot know the impact of selectivity ex ante. The selectivity correction terms must be constructed and included in the estimation of the utility-difference parameters whenever participation might be endogenous. Alternately, joint probit models should be used.

[19] These results, however, contrast with the implications from a more-naive sequentially estimated model that does not recognize in the second stage the estimated nature of the selectivity correction terms created in the first stage.

cannot be ruled out universally in models of this type because it does not appear to be statistically significant in this particular example. It will have to be assessed in every application.

Appendix II outlines some very interesting preliminary findings concerning the robustness of the indirect utility function parameters inferred from responses to three different contingent valuation questions posed on our survey. In particular, it is not possible to reject the equality of the corresponding preference parameters estimated using three entirely different questions.

c.) Empirical Cost-Benefit Quantities

For cost-benefit purposes, we are less interested in the bivariate probit coefficients (or in the two sets of ordinary probit parameter estimates from the sequential model) and more interested in particular functions of these estimated parameters and the data. For all of these quantities, we employ the means for each individual across a set of Krinsky-Robb simulations. Table 4 presents comprehensive descriptive statistics for the marginal distributions (across our entire sample of 4319 respondents) of the individual average cost-benefit quantities. The descriptive statistics including mean, standard deviation, minimum, maximum, skewness and kurtosis. The underlying individual averages are calculated from 50 Krinsky-Robb simulations. Recall that we are treating these individual simulation averages as analogous to point estimates of the individual values, had analytical formulas for their expectations been tractable.

Our estimated option price (the ex ante measure) for acid rain damage to 20% of all currently fishable high altitude lakes is roughly $253. The

Table 4

Descriptive Statistics across Estimating Sample (n = 4319) for Means (over 50 Krinsky-Robb simulations) of Selected Cost-Benefit Quantities (using bivariate probit parameter estimates)

| Measure | Mean (Std. Dev.) | Minimum Maximum | Skewness Kurtosis |
|---|---|---|---|
| $S_n$ (non-user surplus[a]) | \$ 234.62 (200.45) | \$ 13.74 3069.68 | 4.51 41.38 |
| $S_u$ (user surplus[b]) | 375.92 (304.56) | 31.39 4407.29 | 4.32 37.53 |
| E[S] (expected surplus) | 258.16 (217.48) | 14.46 3076.81 | 4.51 41.00 |
| OP (option price) | 252.89 (212.97) | 14.26 3074.16 | 4.51 41.06 |
| OV (option value) | -5.27 (7.73) | -121.63 -0.000409 | -4.047 36.13 |
| E[fb] (expected fair bet) | 288.53[c] (244.24) | 16.82 3259.89 | 4.46 39.85 |

a Fitted values for non-users; simulated values for users. Interpreted as "existence and bequest value" component of demand for all respondents.

b Fitted values for users; simulated values for non-users. Differences $S_u$ - $S_n$ are the additional value that comes with use.

c Interior fair bet points occurred for only 24.5% of the Krinsky-Robb simulations, on average. At most, 72% were interior, but for some individuals, none were. This stems from the near-linearity of the individual WTP loci.

ex post measure, expected consumer surplus, is very little different at about $258. The expected value of the fair bet, however, is approximately $289. These quantitative results are dependent upon the definition of a "user" adopted in the estimation process. These particular values focus on anglers who target trout and who have been fishing this season. (We might, instead, consider anyone who has ever been fishing as a "user," so that participation characterizes just over 80% of the sample. The resulting option price is just less than $280, expected consumer surplus is just over $280, $S_n$ about $158 and $S_u$ about $306.)

For each individual in our sample, our estimate of $S_n$ can be translated as ex post <u>non-user</u> equivalent variation and $S_u$ is <u>total</u> ex post <u>user</u> equivalent variation. Each individual's WTP locus passes through the point S, but its shape depends crucially on the individual's perceived participation probabilities $P_n$ and $P_u$ (see equation (24)). What is the effect of reducing someone's use probability to zero?

For any individual, as $P_u$ approaches zero and $P_n$ approaches one, their WTP locus becomes steeper and steeper (see equations (24-26)) until, in the limit, its intercept with the $\gamma_n$ axis in Figure 1 goes to $S_n$. So $S_n$ is the value each individual would place on the resource if we chose to simulate circumstances wherein it was impossible for him or her to use the resource. We interpret $S_n$ as "existence and/or bequest value." By similar reasoning, $S_u - S_n$ can be considered the incremental value of the resource that comes with use. This analysis, therefore, suggests that non-use values are a very substantial component of total user value.

Assuming diminishing marginal utility for fishable lakes, the sample average $S_n$ value, $235, can be interpreted as an estimate of the per-capita existence value of "the last" 20% of all currently fishable high-altitude lakes. Five times this dollar figure might, therefore, be considered a lower bound on the per-capita existence value of the fishability of all high-altitude lakes in the Northeast. If the represented population consists of roughly 13 million people, a crude estimate of the total dollar value is $15.3 billion annually ($235 * 5 * 13 million).

Our sample median individual option price is roughly $213, indicating that a majority of our sample would be willing to vote in favor of a policy to prevent the damage that resulted in uniform costs of $213 to each person. The types of individuals who are likely to be willing to pay more, or less, are also revealed by our model. This information could be very useful to either the government or to industry for identifying and mobilizing their respective constituencies for a referendum on this issue.[20]

Our analysis has focused upon option prices under the assumption that only a common "sure" payment for acid rain mitigation would be implemented, despite identifiable differences in user/non-user contingent payments for each individual. In practice it is possible to go part way towards usage-contingent payments with a crude mixed payment strategy. The uniform component might consist of cost to everyone (for example, via higher product prices) equal to the mean existence and bequest value. Users could then be "taxed" (through user fees) an amount necessary to match the total use/non-use differential for current users. This tax would probably have to be larger than the current average use/non-use differential because of the

[20] The lower quartile of option prices in our sample is $125 and the upper quartile is $318. A thorough analysis of the sensitivity of value estimates to the presence of influential observations would perhaps include a re-estimation of the model using only respondents having option prices, say, in the inter-decile range according to this basic model.

effect of such a tax on participation. In a somewhat richer specification, the participation elasticity with respect to annual fees can be approximated from a participation model. At present, omitted variables leave income with an apparent negative effect on participation, so this payment scenario cannot be simulated reliably.

As was mentioned in the discussion of the stochastic specification, the individual WTP loci in this example are almost linear, having little curvature because (Y-S)/Y is so close to unity. Plots of the WTP locus for every individual could readily be generated, and there is great variability in the slopes and locations of these curves across respondents. Option price, expected consumer surplus, and option value can easily be computed for each individual in the sample.

The aggregate WTP locus is not relevant in this application because the points from which it is assembled are determined under the assumption of identical "community" uncertainty ($P_n/P_u$) for all individuals. Individual probabilities of use and non-use will always differ. The only time they would be identical would be if the resource was eliminated so that all probabilities fell to zero.

d.) Using the Calibrated Model for Forecasting and Simulation

The estimated structural parameters for freshwater fishing participation probabilities and for the indirect utility difference function associated with our acid rain CV question are interesting in their own right. So are the estimates of current cost-benefit quantities: individual option prices, expected surpluses, option values, and individual WTP loci. However, our model is also particularly well-suited for forecasting and simulation because we have been careful to control for a wide range of factors which can influence participation and resource values.

In the most general case, we can consider a unilateral permutation of a variable that appears both among the Z variables in the participation probability model and among the X variables in the CV indirect utility model. For example, the age distribution in the US is expected to shift upwards over the next twenty years due to the "graying" of the baby-boom generation. By 2010, the average U.S. citizen will be roughly 5 years older. We can retain the calibrated parameters of our participation submodel and our main CV indirect utility difference model and counterfactually simulate a crude approximation to this change by arbitrarily adding 5 years to the age of every respondent in the sample.[21]

The shift in the age distribution will change fitted participation probabilities in each of the five subsequent years due to the dynamic nature of our participation model. These altered participation probabilities will directly affect three components of the model: the slope of the iso-expected value line facing each individual, the location of each individual's S point, and the shape of each individual's WTP locus. But each person's S point and WTP locus are also determined by point estimates produced by the second-stage CV indirect utility function model, and a permutation of age for each person will change all of the estimated $f_j$ quantities as well. Figure 2 illustrates the flow of effects from permuting the values of the age variable.

Rather than detailing the effects of our simulated age change at each

[21] Our cross-sectional data set does not allow this important distinction between age effects and cohort effects. Information from a panel-type survey would be required for this. This simulation therefore ignores cohort effects but still illustrates the forecasting technique.

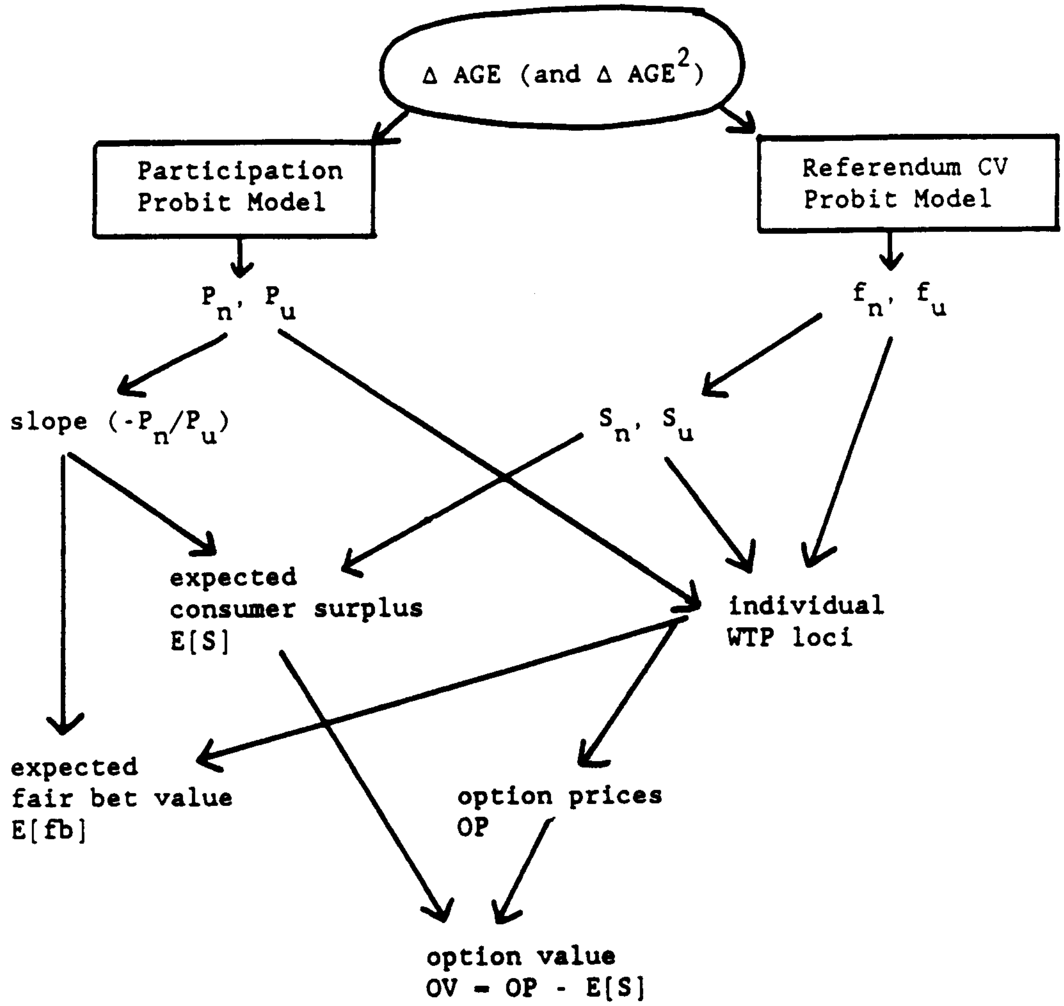


Figure 2: Pathways for Effects of Counterfactual Simulations

stage in Figure 2, we summarize the overall effect in Table 5 by showing how the distributions of each of the candidate welfare measures change when all ages are permuted upwards by 5 years (Simulation 1). The effects are very little different if we additionally force into retirement (RETI=1) all respondents older than 65 after this age change (Simulation 2). Simulation 3 is a $4000 increase in real income for all respondents. This real income increase is also approximately consistent with forecasts for the year 2010. The separate effects of the age increase and the income increase are partially offsetting. If only the age increase took place, acid rain damage prevention would be a potential Pareto improvement if per-capita costs were less than about $241 (instead of $253). If only the real income increase took place, prevention would remain potentially Pareto-improving even if costs were as high as than $280.

It is important to note that our model makes it perfectly feasible to simulate simultaneous shifts in any or all of the exogenous variables which enter our participation and/or indirect utility-difference models. The overall effects of these simultaneous changes can then be assessed using any of the re-calculated cost-benefit quantities or the new WTP loci.

## 6. CONCLUSIONS

This paper develops a utility-theoretic empirical implementation of Graham's (1981) application of state preference theory to cost-benefit analysis under uncertainty. Respondents' revealed behavior is used to infer their likely subjective participation probabilities. This information is then combined with a basic indirect utility function calibrated using responses to a referendum contingent valuation question concerning acid rain

Table 5

Descriptive Statistics across Estimating Sample (n = 4319)
for Means (over 50 Krinsky-Robb simulations)
of Selected Cost-Benefit Quantities

| Measure | FITTED VALUES<br><br>Mean<br>(Std.Dev.) | SIMULATION 1<br>Increase AGE by 5 years:<br>Mean<br>(Std.Dev.) | SIMULATION 2<br>Increase AGE and impose retirement[a]<br>Mean<br>(Std.Dev.) | SIMULATION 3<br>Increase Y by $4000:<br>Mean<br>(Std.Dev.) |
|---|---|---|---|---|
| $S_n$ (non-user surplus[b]) | $ 234.62 (200.45) | $ 225.88 (193.63) | $ 224.81 (193.19) | $ 259.75 (202.71) |
| $S_u$ (user surplus) | 375.92 (304.56) | 370.12 (300.36) | 366.98 (298.95) | 416.82 (306.86) |
| E[S] (expected surplus) | 258.16 (217.48) | 246.43 (208.14) | 245.27 (207.75) | 285.99 (220.04) |
| OP (option price) | 252.89 (212.97) | 241.57 (204.04) | 240.44 (203.64) | 280.13 (215.41) |
| OV (option value) | -5.27 (7.73) | -4.86 (7.49) | -4.83 (7.46) | -5.86 (8.19) |
| E[fb] (expected fair bet) | 288.53 (244.24) | 273.82[c] (232.88) | 272.69 (232.57) | 319.65[d] (247.27) |

[a] Retire everybody who is 65 or older (after the simulated age increase).

[b] As in Table 4, interpreted as "existence and bequest value."

[c] Interior fair bet points occurred for only 26.6% of the Krinsky-Robb simulations, on average. At most, 72% were interior; in some cases, none were. (Very similar results for Simulation 2.)

[d] Interior fair bet points occurred for only 24.6% of the Krinsky-Robb simulations, on average. At most, 72% were interior; in some cases, none were.

damage to freshwater resources in the Northeast United States. Participation probabilities are specifically recognized to be endogenous in our model of individual resource values. We are able to attain most of the interesting alternative measures commonly considered in cost-benefit analyses in the presence of uncertainty.

As was shown by Graham (1981) and recently reiterated by Smith (1990), in the absence of individually insurable risks, the best measure of individual welfare for ex ante cost-benefit analysis is the option price associated with the project for each individual. This measure is unambiguously identified in our model and is very easy to calculate, as are other measures developed in the theoretical literature and relevant to our application (with the sole exception of the "certainty point"). Our model even allows different welfare concepts to be used for different groups of individuals in a sample, if so desired, since all measures are available for everyone.

Several issues keep re-emerging in the literature on cost-benefit analysis under uncertainty: (i.) how do changes in the likelihood of participation change the value of the good? and (ii.) how does the existence or nonexistence of actuarially fair insurance markets affect the social value of a project (or, here, an environmental change)? (iii.) what about option value? and (iv.) how should existence values be measured empirically?

The first issue can be addressed in a perfectly straightforward manner in the framework of our model. The calibrated first-stage probability model and second-stage indirect utility model can readily accept counterfactually simulated changes in any of the wide range of variables which drive the model. These simulations produce new probabilities and new individual willingness-to-pay loci for each respondent in the sample. New values for the alternative welfare measures produced by our model are then available.

Since endogenous participation probabilities are explicitly modelled in this study, we can examine the factors which influence demand uncertainty. To our knowledge, no other empirical study had yet adopted this strategy. Purely exogenous changes in participation probabilities can also be readily simulated. If some purely arbitrary factor is expected to change only the individual probabilities without affecting the utility function(s), these new probabilities can also be "cranked" through the model to provide our full array of valuable information about anticipated welfare consequences. Factors which uniformly influence community risk, as opposed to individual risk, could also be modelled in this fashion.

The second issue concerns welfare measurement when individuals are, or are not, able to obtain fair insurance against their uncertainty regarding whether or not they will be participants. Without actuarily fair insurance for individuals against their user/non-user status, no expected value measure of welfare is relevant (neither the expected surplus, E[S], nor the expected value of the fair bet point, E[fb]). If such insurance is available, the fair bet point has the largest expected value of all pairs of contingent payments along the individual WTP locus and, as such, it conveys the most appropriate measure of benefits. The fair bet point is easily solved-for from each individual's willingness-to-pay locus in our empirical model.

Our response to the third issue, option value, succinctly summarizes one of the main contributions of this paper. The considerable debate over the sign and size of option value has been motivated by the need to quantify the difference between the first-best and second-best measures of welfare. The earlier work was based on the premise that first-best measures often cannot be estimated. The development of a modelling strategy and an empirical technique which allows researchers to directly estimate first-best measures has been the goal in this research. By estimating the parameters of a well-defined indirect utility-difference function and deriving the corresponding individual willingness-to-pay loci, each of the potentially first-best measures of welfare become available to the researcher.

Finally, our methodology illustrates an intuitively appealing derivation of existence and bequest (non-use) values for environmental resources. Non-use value is, on the one hand, value accruing to non-users, but the mean value to this group may be an inappropriate statistic because of self-selection into this group. On the other hand, non-use values also accrue to current users of the resource, so that total value must be decomposed into non-use value and the increment to value that comes with use. In order to achieve this decomposition, it is necessary to be able to simulate, for users, what their individual values would be if they had been non-users. Our framework readily accommodates this task.

There remain several issues for future research to address. One issue is the recursive nature of the current model. While the two step process of separating recreation behavior into the participation decision and then demand is common, a fully joint model would be preferable. Analysis using an open ended contingent valuation survey may prove more amenable to a fully simultaneous model. A second topic for further research is whether this model could be implemented with revealed preference data. A model based upon revealed preference data would provide an interesting alternative to estimates of option price and existence value derived from contingent valuation studies. Finally, the model could be expanded to address multiple states of the world, rather than the simple two-state scenario analyzed in this study.

## REFERENCES

Brookshire, D.S., L.S. Eubanks, and A. Randall (1983) "Estimating Option Prices and Existence Values for Wildlife Resources," Land Economics (59) 1-15.

Brownstone, D. (1991) "Multiple Imputations for Dynamic Econometric Models," working paper, Department of Economics, U.C. Irvine.

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

Cameron, T.A. and M.D. James (1987) "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys," Review of Economics and Statistics (69) 269-276.

Chavas, J.-P. (1991) "On Welfare Analysis under Temporal Uncertainty," Land Economics (67) 37-48.

Chavas, J.-P., R.C. Bishop, and K. Segerson (1986) "Ex Ante Consumer Welfare Evaluation in Cost-Benefit Analysis," Journal of Environmental Economics and Management (13) 255-68.

Chipman, J.S. and J.C. Moore (1980) "Compensating Variation, Consumer Surplus and Welfare," American Economic Review (70) 933-49.

Colby, B.G., and D.C. Cory (1989) "Valuing Amenity Resources under Uncertainty: Does the Existence of Contingent Claims Markets Matter?" Journal of Environmental Economics and Management (16) 149-55.

Cory, D.C. and B.C. Saliba (1987) "Requiem for Option Value," Land Economics (63), 1-10.

Cummings, R.G., D.S. Brookshire, and W.D. Schulze (1986), Valuing Environmental Goods: An Assessment of the Contingent Valuation Method, Totowa, New Jersey: Rowman and Allenheld.

Goldfeld, S., and R.E. Quandt (1986) GOOPT. Princeton, NJ: Princeton University.

Graham, D.A. (1981) "Cost-Benefit Analysis under Uncertainty," American Economic Review (71) 715-725.

Graham, D.A. (1984) " Cost-Benefit Analysis under Uncertainty: Reply" American Economic Review (74) 1100-1102.

Greene, W.H. (1990) Econometric Analysis. New York: Macmillan.

Greenley, D.A., R.G. Walsh, and R.A. Young (1981) "Option Value: Empirical

Evidence From a Case Study of Recreation and Water Quality," Quarterly Journal of Economics (96) 657-683.

Hammond, P.J. (1981) "Ex-Ante and Ex-Post Welfare Optimality Under Uncertainty," Economica (48) 235-50.

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

Hausman, J.A. (1981) "Exact Consumer Surplus and Deadweight Loss," American Economic Review (71), 662-76.

Hirshleifer, J. (1965) "Investment Decision under Uncertainty: Choice-Theoretic Approaches," Quarterly Journal of Economics (79), 509-536.

Hirshleifer, J. (1966) "Investment Decision under Uncertainty: Applications of the State-Preference Approach," Quarterly Journal of Economics (80), 252-275.

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

Krutilla, J.V. (1967) "Conservation Reconsidered," American Economic Review (57) 777-786.

Madariaga, B., and K.E. McConnell (1987) "Exploring Existence Value," Water Resources Research (23) 936-942.

Maddala, G.S. (1983) Limited-dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

McConnell, K.E. (1983) "Existence Value and Bequest Value," in R.D.Rowe and L.G. Chestnut, eds., Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Boulder, CO: Westview Press.

McKenzie, G.W. and I.F. Pearce (1982) "Welfare Measurement: A Synthesis," American Economic Review (72) 669-82.

Mendelsohn, R. and W.J. Strang (1984) "Cost-Benefit Analysis under Uncertainty: Comment," American Economic Review, (74) 1096-1099.

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

Randall, A. and J.R. Stoll (1983) "Existence Value and a Total Valuation Framework," in R.D. Rowe and L.G. Chestnut, eds., Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Boulder, Co: Westview Press.

Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys, New York: Wiley.

Schmalensee, R. (1972) "Option Value and Consumer Surplus: Valuing Price Changes Under Uncertainty," American Economic Review (62), 813-24.

Shankle, S.A., J.E. Englin, W.H. Monroe, D.M. Beck, M.F. Casinelli, H. Saunders, and S.E. Grover, (1990) "Freshwater-Based Recreation in the Upper Northeast: Data Collection Documentation," Battelle Memorial Institute, Pacific Northwest Laboratory, USDOE, prepared for National Acid Precipitation Assessment Program, November.

Smith, V.K. (1987a) "Uncertainty, Benefit-Cost Analysis and the Treatment of Option Value," Journal of Environmental Economics and Management (14) 283-292.

Smith, V.K. (1987b) "Non-Use Values in Benefit-Cost Analysis," Southern Economic Journal (54) 19-26.

Smith, V.K. (1990) "Valuing Amenity Resources under Uncertainty: A Skeptical View of Recent Resolutions," Journal of Environmental Economics and Management (19) 193-202.

Walsh, R.G., J.B. Loomis, and R.A. Gillman (1984) "Valuing Option, Existence, and Bequest Demands for Wilderness," Land Economics (60) 14-29.

Weisbrod, B. (1964) "Collective-Consumption Services of Individual-Consumption Goods," Quarterly Journal of Economics (78) 471-77.

Willig, R.D. (1976) "Consumer's Surplus without Apology," American Economic Review (66) 589-97.

# APPENDIX I

## Distillation of the Survey Data

a.) Geographical Information

The forty randomly selected counties belong to individual states as follows: counties 1 to 6 (Maine), counties 7 to 10 (New Hampshire), counties 11 to 34 (New York), and counties 35 to 40 (Vermont).

Maps of the Northeast were used to determine whether each county contains or is adjacent to a major urban area. This information is captured by the rough dummy variable URB.

Counties were matched with their corresponding so-called FIPS codes so that data could be spliced in from the NORSIS data files.

If the age of the respondent is unavailable, observations were dropped from the data set.

If the person has taken a freshwater fishing trip in the relevant area since April 1 of the current year, the participation dummy variable is set equal to 1. If not, the dummy is set equal to zero. An alternative assumption also assumes current participation if any trips are planned between now and the end of the current year.

The contingent valuation questions are all of the "closed-ended" or "referendum" contingent valuation variety. Possible responses include: YES, NO, DON'T KNOW, and REFUSED. Initial analyses will focus on the YES and NO responses, but attention will also be devoted to the determinants of the ambiguous responses. (Furthermore, interviewers were asked to circle question numbers which they believed that the respondent had difficulty answering. This additional information could also potentially be utilized at a later date.)

The dollar amounts entered on the questionnaires were randomly assigned in advance from a predetermined list of values between $1 and $100. The full set of CV questions was:

Q-5: "If the abundance of trout could be doubled in lakes and streams in the Northeast, would you be willing to pay ____ per year for this benefit?"

Q-6: "If algae growth from sources such as agricultural runoff could be reduced so that you could see twice as deep into lakes and rivers would you be willing to pay ____ per year for this benefit?"

Q-7: "If acid rain damaged fishing in 20% of all currently fishable high altitude lakes in the Northeast, would you be willing to pay ____ per year to prevent this?"

Q-8: "Suppose half of all lakes in the Northeast were unfit for swimming due to pollution would you be willing to pay ____ per year to prevent this?"

If the respondent owns a secondary residence, the dummy variable SECR equals one, otherwise it is zero.

The available ten categories of educational attainment (highest level of education completed) are coded as a smaller set of six dummy variables. The omitted category is "High School Graduate or less". The omitted category includes individuals with some college education. COLG is "College Graduate or postgraduate education." TRSC is trade school.

The respondent's current employment situation is summarized in six categories. The omitted category is "Employed Full Time." Other dummy variables are PART (Employed Part Time), NOEM (Not Employed Outside the Home), RETI (Retired), STUD (Student), and UNEM (Unemployed).

Racial and ethnic background information was also solicited from respondents. The omitted category is WHITE. Dummy variables indicated American Indian (AMIN), Mexican American (MEXA), Asian American (ASIA), and Black (BLK).

As is typical in these surveys, income data are only available at a very coarse level. The household's total income before taxes in 1988 was requested and intervals were implied by the respondent's answers. Representative values (MIDINC) were assigned somewhat arbitrarily as follows:

| | | |
|---|---|---|
| less than $ 10,000 | coded as 7 | ($ '000) |
| $ 10,001 to $ 20,000 | 15 | |
| $ 20,001 to $ 30,000 | 25 | |
| $ 30,001 to $ 40,000 | 35 | |
| $ 40,001 to $ 50,000 | 45 | |
| $ 50,001 to $ 60,000 | 55 | |
| $ 60,001 to $ 70,000 | 65 | |
| $ 70,001 to $ 80,000 | 75 | |
| $ 80,001 to $ 90,000 | 85 | |
| $ 90,001 to $ 100,000 | 95 | |
| $ 100,000 to $ 250,000 | 150 | |
| more than $ 250,000 | 300 | |

For the comprehensive error bars required for the ultimate empirical implications of a study based on these data, sensitivity analysis with respect to these representative values has been conducted.

The gender dummy variable is FEM, equal to one for female respondents and zero for males.

For the initial estimates, we deleted respondents with incomplete data for any of the prospectively important explanatory variables. Observations were deleted if there were no data for the original variables: SECR, UNEM, PART, NOEM, RETI, STUD, GRSC, SOMH, TRSC, COLG, MAST, DOCT, URB, AGE, AMIN, ASIA, BLK, MIDINC, and FEM.

Some 285 variables are available, by county FIPS code, in the so-called NORSIS data set. The level of disaggregation of these variables is astonishing, and previous users have some reservations about the reliability of some of these data. However, it is helpful to extract some basic information from this data set on the relative availability of water-based recreational opportunities within each county.

The variables presently used in the models are:

POP - county population estimates for 1985 (census 1984)
CNTY - total county area (census)
LILL - Acres of water in bodies <2 acres
LILR - acres of water in rivers/streams <66 feet wide

## APPENDIX II

### Robustness of the Utility Parameter Estimates

As we have noted in the body of the paper, we had responses to three other contingent valuation questions in addition to the acid rain question. We initially estimated entirely separate indirect utility difference function parameters for each of these questions. One of the other questions concerned a reduction in algae growth that would allow visibility to twice current depths (amenity level from A to 2A); a second concerned controlling pollution that would prevent swimming in half of all lakes (amenity level from A to .5A).[22] Across these three models, an interesting result emerged. Our point estimates of the marginal utility of log(income) to users and to non-users were virtually identical for the three cases!

By estimating the indirect utility difference separately for each question, we were allowing several different point estimates for what ought to be the same utility function parameters. This observation suggested that all three contingent valuation scenarios could be embedded in one model. Equations (10) and (11) in the body of the paper can be replaced by a set of four extended equations:

$$(31) \qquad V_j^0 = \beta_j \log(Y) \quad + \delta_j \log P + \alpha_j^{0\prime} X$$
$$+ \omega_{j1}\log(A_1) + \omega_{j2}\log(.8A_2) + \omega_{j3}\log(.5A_3) + \eta^0, \qquad j = n, u$$

$$(32) \qquad V_j^{11} = \beta_j \log(Y-t_1) \quad + \delta_j \log P + \alpha_j^{1\prime} X$$
$$+ \omega_{j1}\log(2A_1) + \omega_{j2}\log(A_2) + \omega_{j3}\log(A_3) + \eta^{11}, \qquad j = n, u$$

$$(33) \qquad V_j^{12} = \beta_j \log(Y-t_2) \quad + \delta_j \log P + \alpha_j^{1\prime} X$$
$$+ \omega_{j1}\log(A_1) + \omega_{j2}\log(A_2) + \omega_{j3}\log(A_3) + \eta^{12}, \qquad j = n, u$$

$$(34) \qquad V_j^{13} = \beta_j \log(Y-t_3) \quad + \delta_j \log P + \alpha_j^{1} X$$
$$+ \omega_{j1}\log(A_1) + \omega_{j2}\log(A_2) + \omega_{j3}\log(A_3) + \eta^{13}, \qquad j = n, u$$

---

[22] For one question, regarding WTP for doubling the abundance of trout, fitted values were extremely tiny relative to income (on the order of $10), and we have deemed the estimates for this question to be unreliable.

where $A_1$, $A_2$, and $A_3$ are now three distinct environmental amenities. The corresponding indirect utility-difference functions to be estimated using probit techniques become:

$$(35) \qquad V_j^{11} - V_j^0 = \beta_j \log[(Y-t)/Y] + \alpha_j'X + \omega_{j1}\log(2) + \rho_n \lambda^n + \rho_u \lambda^u + \eta'^1$$

$$(36) \qquad V_j^{12} - V_j^0 = \beta_j \log[(Y-t)/Y] + \alpha_j'X - \omega_{j2}\log(.8) + \rho_n \lambda^n + \rho_u \lambda^u + \eta'^2$$

$$(37) \qquad V_j^{13} - V_j^0 = \beta_j \log[(Y-t)/Y] + \alpha_j'X - \omega_{j3}\log(.5) + \rho_n \lambda^n + \rho_u \lambda^u + \eta'^3$$

where $j = n, u$ in all three equations. This means that the probit intercept term among the $\alpha_j$ coefficients will have a different constant term added to it depending upon which question is being analyzed and upon the user/non-user status of the respondent. We cannot identify $\omega_{j1}\log(2)$, but we can estimate the difference between $\omega_{j1}\log(2)$ and $\omega_{j2}\log(.8)$, and the difference between $\omega_{j1}\log(2)$ and $\omega_{j3}\log(.5)$ by appropriate use of question-indicator dummy variables with the pooled data.

If we are willing to assume that $\eta'^1$, $\eta'^2$ and $\eta'^3$ have independent and identical normal distributions, each answer in a person's set of three responses to the CV questions can then be treated as a separate observation. This gives us three glimpses of their indirect utility function. For the exploratory models we have examined, we chose a subset of respondents who answered all three CV questions, yielding a sample of size 3349. Constraining the corresponding coefficients to be the same across all three responses involves stacking the data set to produce (3*3349) = 10,047 "observations."[23]

For our pooled model, we conform the specifications by creating a set of X variables (individual attributes) consisting of the union of all X variables previously employed for the analysis of each individual contingent valuation question. We estimated three separate models, attaining individual maximized log-likelihood values of -1980.9, -1794.8 and -1507.2. When summed, the implied log-likelihood across all three questions for the pooled data and no restrictions is -5282.9. Pooling the data involves 78 parameter restrictions and produces a maximized log-likelihood of -5309.9. The value of the LR test statistic is therefore 54.0. The 5% critical value of a $\chi^2(78)$ is on the order of 100, so we cannot reject the restrictions.

Another important test concerns the necessity of allowing all of the

---

[23] Of course, there may be unique individual fixed effects that could be recognized in a more elaborate specification. However, panel data techniques in the case of probit models are not as straightforward as they are in OLS models with continuous dependent variables.

parameters of the indirect utility-difference function to vary across users and non-users of the resource. One obvious alternative that would substantially reduce the dimensionality of the parameter space would be to suppress all of the user-dummy interaction variables to leave only an intercept shift dummy for users. This hypothesis involves 17 restrictions and produces a likelihood ratio test statistic of 44.8. The 5% critical value in this case is only 27.58, so we reject these restrictions.

If we were to pursue more elaborate specifications, we would clearly choose to rely upon the indirect utility parameters and the resulting cost-benefit quantities derived from the pooled data on all three questions. However, since our mission in this paper is to illustrate the empirical implementation of Graham's theory, we will use the inferences from the pooled data model merely to support the robustness of our utility-difference parameter estimates.