

Combining Sources of Data to Estimate Preferences for an Environmental Resource

Edward R. Morey and William S. Breffle
Department of Economics
Campus Box 256
University of Colorado
Boulder CO 80309-0256
contact: edward.morey@colorado.edu

November 7, 2003

Abstract

The objective of this research is to develop a random-utility model of preferences that is estimated with three types of data: stated preference choice data, stated preference frequency data, and revealed preference use data. A probit model is specified, and the application is angler preferences for Green Bay. The model takes advantage of the relative strengths of the different types of data; the combination of data not only allows for the estimation of how individuals trade off different characteristics of the site, but also how the angler's proportion of trips to Green Bay would change if its characteristics changed. This methodology for multiple data sets is especially useful when the commodity of interest is unique, as is Green Bay. The estimated parameters are used to estimate the compensating variation per fishing day (and per Green Bay fishing day) for the elimination of Green Bay fish consumption advisories.

* This work has benefitted greatly from comments and suggestions from Vic Adamowicz, David Allen, Bob Baumgartner, Rich Bishop, Don Dillman, David Layton, Pam Rathbun, Bob Rowe, Paul Ruud, V. Kerry Smith, Roger Tourangeau, Don Waldman, and Michael Welsh.

1. Introduction

A goal of environmental economics is to explain, estimate, and predict the demand for and value of environmental commodities as a function of their costs and characteristics. These commodities are typically not directly bought and sold in the marketplace; they are often unique in the sense they have few substitutes. Often the intent is to value environmental resources at levels that do not currently exist. If some characteristics of the new commodity are not present in existing commodities, or there is not sufficient variation in the characteristics of existing commodities, estimation will not be possible using only market data. A solution is to use stated preference (SP) data, either by itself or combined with revealed preference (RP) data. We combine RP frequency data with SP choice *and* frequency data.¹

SP and RP data provide different information about preferences, so combining them leads to better estimates of those preferences.² However, doing so raises interesting issues with respect to modeling and estimation. The preference information in different data types takes different forms, and these need to be integrated into a utility-theoretic model.

The objective of this research is to develop a random-utility model that combines these three types of data. A probit model is specified, and the application is angler preferences for Green Bay. Parameter estimates are obtained for two conditional indirect utility functions: one

1. A few environmental applications have used SP frequency data only, such as Adamowicz et al. (1994) and Englin and Cameron (1996), and a multitude of studies across disciplines have SP choice questions, which evolved from conjoint analysis. Cattin and Wittink (1982) and Wittink and Cattin (1989) survey the commercial use of conjoint analysis, which is widespread. For survey articles and reviews of conjoint, see Louviere (1988, 1992), Green and Srinivasan (1990), and Batsell and Louviere (1991). Transportation planners use choice questions to determine how commuters would respond to a new mode of transportation or a change in an existing mode. Hensher (1994) provides an overview of choice questions as they have been applied in transportation. Two recent environmental applications using choice questions are Boxall and Adamowicz (2002) and Breffle and Rowe (2002).

2. This practice is widely supported. See, for example, McFadden (1986), Ben-Akiva and Morikawa (1990), Morikawa et al. (1990), Cameron (1992), Louviere (1992), Hensher and Bradley (1993), Adamowicz et al. (1994, 1997, 1998), Ben-Akiva et al. (1994), Swait et al. (1994), Morikawa et al. (1991), Louviere (1996), Kling (1997), and Mathews et al. (1997).

for a Green Bay fishing day, and one for fishing elsewhere. We do not model the angler's total number of fishing days, only the proportion to Green Bay and how the angler trades off Green Bay characteristics. Not modeling the total number of fishing days simplifies data collection and modeling while still generating policy-relevant results that are often easier to defend than the results of models that explain and predict both participation and site choice.

The estimated conditional indirect utility functions are used to estimate two compensating variation measures for an improvement in Green Bay: compensating variation per Green Bay fishing day, and expected compensating variation per fishing day to all sites. Compensating variation per Green Bay fishing day multiplied by the individual's current number of Green Bay fishing days is a lower-bound estimate of the individual's yearly compensating variation for the improvement, and so is compensating variation per fishing day multiplied by the angler's current number of fishing days (Morey, 1994). The first product is smaller than the second because it restrictively assumes the angler will not increase his proportion of fishing days to Green Bay when Green Bay is improved. Compensating variation per fishing day multiplied by the angler's current number of fishing days incorporates this likely possibility.

The RP data consist of the total number of fishing days to all sites for each individual in the sample, and the number of those days the angler fished Green Bay under current conditions. The SP data include the responses to choice questions. Each sampled individual indicated the choice between a pair of Green Bay alternatives (Green Bay under different conditions). Green Bay is characterized in terms of catch rates and fish consumption advisories (FCA) levels for yellow perch, trout and salmon, walleye, and smallmouth bass, and an angler's share of the daily launch fee. Next, in a follow-up question to each pair, respondents indicated the proportion of

trips they would choose to fish Green Bay with the characteristics described in their chosen alternative. For each sampled individual, these two questions were repeated eight times, where the characteristics of the Green Bay alternatives in the pairs vary over the eight pairs. The use of SP data was deemed to be necessary because Green Bay is a unique fishing site in terms of size and species mix, and inland waters do not have FCAs caused by PCB contamination.

Choice questions encourage respondents to concentrate on the tradeoffs between characteristics. While such questions alone tell nothing about the proportion of trips the angler would take to Green Bay under different conditions, they can be used to determine how much an angler would be willing to pay per Green Bay fishing day to fish Green Bay without FCAs. The choice questions alone also allow the estimation of how much other Green Bay characteristics (e.g., catch rates) would have to increase to make the angler's utility the same as it would be absent FCAs. Adding the frequency data (RP and SP) allows the estimation of how much relative demand for trips to Green Bay would change in the absence of the FCAs.

2. Population, sampling, and response rates

The target population is current Green Bay anglers who live in the area: anglers who purchased licenses in eight counties near Green Bay and who fished Green Bay in 1998. A three-step procedure was used in 1998 to collect data from a random sample of the targeted anglers. First, a random sample of anglers was drawn from 1997 license holders in the county courthouses in the eight targeted counties. Second, using the license holder list, a telephone survey was conducted to identify and recruit Green Bay anglers for a followup mail survey. The telephone survey collected some attitudinal data and data on the number of fishing days at Green Bay and elsewhere. The overall response rate to the telephone survey was 69.4%. Third, a mail survey

with the SP questions was conducted with current Green Bay anglers. The response rate to the mail survey was 78.9%, yielding a data set of 647 individual anglers used in the model.

3. The model

The model is developed here component by component. Section 3.1 specifies the choice probabilities for the two Green Bay alternatives in each of the SP choice questions. Section 3.2 adds the SP frequency data and RP data on the total number of fishing days to all sites under current conditions. Section 3.3 incorporates the final component of the data, the RP data on the total number of fishing days to Green Bay under current conditions.

3.1 *Modeling the answers to the SP Green Bay choice pairs*

The model is a discrete-choice random-utility probit model. The utility individual i gets from a day of fishing Green Bay is assumed to be a function of its cost and the characteristics of Green Bay. Consider a standard conditional indirect utility function for fishing Green Bay, $U_i = V + \varepsilon_i$. Since characteristics will vary across each alternative in each of the choice pairs, subscripts and superscripts are needed to keep track of all the variations. Specifically, let the utility for the Green Bay SP choice question alternatives be given by:

$$U_{ij}^{k_{ij}} = \beta' x_{ij}^{k_{ij}} + \varepsilon_{ij}^{k_{ij}}, \quad i = 1, \dots, m; j = 1, \dots, J; k_{ij} \in [1, 2], \quad (1)$$

where $U_{ij}^{k_{ij}}$ is the utility of the k -th alternative of pair j to individual i . That is, i indexes the m respondents, j indexes the eight pairs, and k_{ij} indicates which of the two alternatives within each pair is chosen. The $L \times 1$ vector $x_{ij}^{k_{ij}}$ contains the characteristics of the alternatives, and hence the elements of the unknown $L \times 1$ vector β can be interpreted as marginal utilities. The first element of $x_{ij}^{k_{ij}}$ is the difference between choice-occasion income for individual i and the cost of alternative k_{ij} , and the model is restricted to one with a constant marginal utility of money, which is the first element of β . The specific form of Equation (1) is

$$U_{ij}^{k_{ij}} = \sum_{l=p,s,w,b} \beta_l c_l^{k_{ij}} + \sum_{q=2}^9 \beta_{FCAq} FCA_q^{k_{ij}} + \beta_y (y_i - TC_i - fee^{k_{ij}}) + \varepsilon_{ij}^{k_{ij}}, \quad (1a)$$

where y_i and TC_i are choice occasion income and travel cost for individual i , and average catch times by species (c_l , $l = perch, salmon/trout, walleye, bass$) are measured as the time (in hours) it takes on average to catch one fish of a particular species (perch, salmon/trout, walleye, bass). For example, the perch catch time is approximately 0.75 hours.³ Nine possible levels of FCAs are represented by a set of eight dummy variables, each representing a certain configuration of fish consumption advisories for the four target species. The FCA levels corresponding to the dummy variables generally increase in severity, so that $FCA_3 = 1$, all others zero, means more (and/or more severe) restrictions than $FCA_2 = 1$.⁴ A value of zero for all of the dummy variables (FCA_2 through FCA_9) means no restrictions (eat as many of all species as desired). The most restrictive level, $FCA_9 = 1$, is a warning not to eat more than one perch meal per month or *any* of the remaining three species at all.

Since y_i and TC_i do not vary by k_{ij} , these variables disappear from the utility difference relevant for estimation.

The individual is assumed to choose alternative k_{ij} with the probability:⁵

$$P(K_{ij} = k_{ij}) = P_{ij}^{k_{ij}} = P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}), \quad (2)$$

where k_{ij} is the observed value of K_{ij} . Assume $\varepsilon_{ij}^{k_{ij}}$ are independent (across i) and identically distributed mean zero normal random variables, uncorrelated with $x_{ij}^{k_{ij}}$, with constant unknown

3. It is expected that the coefficients of the catch times will be negative.

4. The exception is in moving from FCA_4 to FCA_5 and from FCA_5 to FCA_6 with the consumption of some species becoming more restricted and others less restricted. These anomalies show up in the parameter estimates.

5. In this notation, if the individual chooses alternative $K_{ij} = 1$ [or 2], then the alternative that was not chosen is $3 - K_{ij} = 2$ [or 1].

variance σ_ε^2 . Note that while the $\varepsilon_{ij}^{k_{ij}}$ vary with j , we do not assume the $\varepsilon_{ij}^{k_{ij}}$ are independent across pairs for a given individual.

From Equations 1 and 2, the probability of choosing alternative k_{ij} is:

$$\begin{aligned}
 P_{ij}^{k_{ij}} &= P(\beta' x_{ij}^{k_{ij}} + \varepsilon_{ij}^{k_{ij}} > \beta' x_{ij}^{3-k_{ij}} + \varepsilon_{ij}^{3-k_{ij}}) \\
 &= P[\varepsilon_{ij}^{3-k_{ij}} - \varepsilon_{ij}^{k_{ij}} < -\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})] \\
 &= \Phi[-\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2}\sigma_\varepsilon]
 \end{aligned} \tag{3}$$

where $\sqrt{2}\sigma_\varepsilon$ is the standard deviation of $\varepsilon_{ij}^{3-k_{ij}} - \varepsilon_{ij}^{k_{ij}}$ under assumption 1 and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function. Equation 3 is a standard probit probability function. Equation (3) enters the likelihood function in Section 3.3. The parameter vector β is identified only up to the scale factor $\sqrt{2}\sigma_\varepsilon$, and σ_ε is not identified, since only the sign and not the scale of the dependent variable (the utility difference) is observed.

Nevertheless, we have chosen to list the parameters of the likelihood function $(\beta, \sigma_\varepsilon)$ separately for Equation 3.

3.2 *Frequency of selecting the preferred Green Bay alternative versus another site*

Now consider, in addition to the data on k_{ij} , the responses to the SP frequency question. Assume utility from fishing at another site, elsewhere, is

$$U_{ij}^0 = \gamma_0 + \varepsilon_{ij}^0. \tag{4}$$

There were no data on the characteristics of the alternative fishing sites for the respondents, so utility for the non-Green Bay alternative site is assumed to be constant across individuals and choice occasions, with an additive random disturbance. This is obviously a simplifying assumption; catch time, travel cost, and any fish advisories at other sites (but not income, as it will drop out) are grouped into the error term. Although a component of travel cost such as distance to the site cannot contribute to a utility difference when the site is Green Bay for both

choices, as it is in the binary choice SP data, it could affect the utility differences between other sites and Green Bay. We assume here that the variation in distance to anglers' other sites is not great across anglers. We further assume that any variation is likely to be randomly distributed across anglers (anglers living close to and far from Green Bay have alternative sites both near and far). The lack of these data adds noise (in the form of increased variance of the disturbance term in Equation 4), but does not bias parameter estimates. Assume the ε_{ij}^0 are independent (across i) and identically distributed normal random variables, with zero expectation and variance σ_0^2 , and $E(\varepsilon_{ij}^0 \varepsilon_{ij}^{k_{ij}}) = \sigma_{\varepsilon 0}$.

Let n_i denote the number of days individual i fished in 1998 and let n_{ij} denote his stated number of days to Green Bay under the conditions in the alternative chosen. The n_{ij} vary from zero to n_i . A plausible assumption is that the n_{ij} are distributed binomially, $N_{ij} \sim B(n_i, p_{ij}^0)$, with probability mass function (conditional on the choice of k_{ij}):

$$P(n_{ij} | K_{ij} = k_{ij}) = \binom{n_i}{n_{ij}} (p_{ij}^0)^{n_{ij}} (1 - p_{ij}^0)^{n_i - n_{ij}}, \quad (5)$$

The parameter p_{ij}^0 in Equation 5 is the probability of choosing Green Bay alternative k_{ij} over the "other" site, conditional on choosing alternative k_{ij} over alternative 3 - k_{ij} :

$$\begin{aligned} p_{ij}^0 &= P(U_{ij}^{k_{ij}} > U_{ij}^0 | U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}) \\ &= P[\varepsilon_{ij}^0 - \varepsilon_{ij}^{k_{ij}} < -\beta'(x_{ij}^0 - x_{ij}^{k_{ij}}) | \varepsilon_{ij}^{3-k_{ij}} - \varepsilon_{ij}^{k_{ij}} < -\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})] \\ &= \frac{\Phi_2[-\beta'(x_{ij}^0 - x_{ij}^{k_{ij}}) / \sigma_{0-\varepsilon}, -\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2} \sigma_{\varepsilon}; \rho]}{\Phi[-\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2} \sigma_{\varepsilon}]} \end{aligned} \quad (6)$$

where $\sigma_{0-\varepsilon}^2 = \text{Var}(\varepsilon_{ij}^0 - \varepsilon_{ij}^{k_{ij}}) = \sigma_0^2 + \sigma_{\varepsilon}^2 - 2\sigma_{\varepsilon 0}$ and where ρ is the correlation between

$$\varepsilon_{ij}^0 - \varepsilon_{ij}^{k_{ij}} \text{ and } \varepsilon_{ij}^{3-k_{ij}} - \varepsilon_{ij}^{k_{ij}},$$

$$\rho = \frac{\sigma_\varepsilon^2}{\sqrt{2\sigma_\varepsilon^2\sigma_{0-\varepsilon}^2}} \quad (7)$$

and Φ and Φ_2 are the standard univariate and bivariate normal distribution functions, respectively (for details of the derivation of Equation 6, see the appendix).⁶

3.3 Incorporating the RP data on actual Green Bay fishing days

In addition to the SP data and the n_i , we have for each i the number of fishing days actually taken to Green Bay, n_i^G (taken, of course, under current conditions). This RP data may be used with the other data in the estimation of the model parameters. Utility for the d -th actual Green Bay fishing day is

given by:

$$U_{id}^G = \sum_{l=p,s,w,b} \beta_l c_l + \sum_{q=2}^9 \beta_{FCAq} FCA_q + \beta_y (y_i - TC_i - fee) + \varepsilon_{id}^G, i = 1, \dots, m, \quad (8)$$

where the values for explanatory variables are the current conditions. Assume the ε_{id}^G are independent (across i) and identically distributed normal random variables, with zero expectation and variance σ_G^2 , and $E(\varepsilon_{id}^G \varepsilon_{ij}^{k_{ij}}) = \sigma_{\varepsilon G}$.

In deciding how many of his fishing days to fish Green Bay, the individual compares utility at Green Bay to utility at other sites, so that the probability of going to Green Bay on day d is:

$$\begin{aligned} P_i^G &= P(U_{id}^G > U_{ij}^0) \\ &= P(\beta'x_i^G + \varepsilon_{id}^G > \beta'x_{ij}^0 + \varepsilon_{ij}^0) \\ &= P(\varepsilon_{ij}^0 - \varepsilon_{id}^G < \beta'x_i^G - \beta'x_{ij}^0) \\ &= \Phi[(\beta'x_i^G - \beta'x_{ij}^0) / \sigma_{0-G}] \end{aligned} \quad (9)$$

6. Note that in Equation 6, β appears twice. On one occasion it is normalized by $\sqrt{2}\sigma_\varepsilon$, and on the other by $\sigma_{0-\varepsilon}$. Also note that under the alternative assumption that disturbances are known to the individual a priori, he would perform the conceptual experiment of generating n_i pairs of disturbances, evaluating utility under the two scenarios, and counting the number of Green Bay trips under the assumption of utility maximization. This process would also imply Equations 5-7.

where:

$$\sigma_{0-G}^2 = \text{Var}(\varepsilon_{ij}^0 - \varepsilon_{ij}^G) = \sigma_0^2 + \sigma_G^2 - 2\sigma_{G0} \quad (10)$$

Since P_i^G is a function of β , the information contained in n_i^G is useful in estimation, and is incorporated into the likelihood. The likelihood function is:

$$L(n_{ij}, k_{ij}, n_i^G, i = 1, \dots, m, j = 1, \dots, J | x_{ij}^1, x_{ij}^2, n_i; \beta, \sigma_0, \sigma_\varepsilon, \sigma_G) = \prod_{i=1}^m \left[\binom{n_i}{n_i^G} (P_i^G)^{n_i^G} (1 - P_i^G)^{n_i - n_i^G} \prod_{j=1}^J P(N_{ij} = n_{ij} | K_{ij} = k_{ij}) P(K_{ij} = k_{ij}) \right] \quad (11)$$

Note that in this likelihood β appears in several expressions: in P_i^G normalized by σ_{0-G} , in $P(N_{ij} = n_{ij} | K_{ij} = k_{ij})$ normalized by $\sigma_{0-\varepsilon}$, and in P_i and $P(K_{ij} = k_{ij})$ normalized by $\sqrt{2}\sigma_\varepsilon$. The ratios of any two of these three parameters are identified in estimation.⁷ The maximum likelihood parameter estimates are consistent. They are also asymptotically efficient under the additional assumption that ε_{ij}^0 and $\varepsilon_{ij}^{k_{ij}}$ are uncorrelated across j .

4. Implementation and estimation

Convergence was achieved for a variety of starting values, and always at the same point. Table 1 provides the estimated values of the parameters and their estimated asymptotic t -statistics. All parameters have the expected signs and all are statistically significant by conventional standards. The perch catch time has by far the largest parameter of the four species' catch times. The pattern of estimated coefficients on the FCA variables is somewhat striking: as the FCA level increases they increase (in absolute value) nearly uniformly, and where they do not, it is as expected (see footnote 4). The same is true for their precision, as measured by their asymptotic t -statistics.

7. Although all parameters are listed separately, it is evident that normalizations are necessary.

The parameters $\sigma_{0-\varepsilon}$ and σ_{0-G} are the standard deviations of the error differences $\varepsilon_{ij}^0 - \varepsilon_{ij}^{k_{ij}}$ and $\varepsilon_{ij}^0 - \varepsilon_{ij}^G$, respectively. Since we have allowed for nonzero covariances between the errors ($\sigma_{\varepsilon 0}$ and σ_{G0} ; see Sections 3.2 and 3.3), identification of individual components of $\sigma_{0-\varepsilon}$ and σ_{0-G} , in particular σ_{ε}^2 and σ_G^2 , is not possible. Thus we are not able to answer the question of which data, RP or SP, contain more information. This question is sometimes addressed in studies that use both RP and SP data by assuming that the disturbances are uncorrelated (see, for example, Ben-Akiva and Morikawa, 1990; Bradley and Daly, 1991; Hensher and Bradley, 1993; Louviere, 1996; and Swait and Louviere, 1993).⁸ For our data and model, this restrictive assumption is easily rejected.

Parameter	Estimate (Asy. <i>t</i> -ratio)
β_y	0.0535 (20.57)
β_p	- 0.5307 (-14.99)
β_t	- 0.0212 (-7.58)
β_w	- 0.0287 (-11.95)
β_b	- 0.0231 (-11.44)
β_{FCA2}	- 0.0972 (-3.07)
β_{FCA3}	- 0.2599 (-7.65)
β_{FCA4}	- 0.5215 (-12.92)
β_{FCA5}	- 0.6017 (-15.80)
β_{FCA6}	- 0.5303 (-13.08)
β_{FCA7}	- 0.7660 (-18.91)
β_{FCA8}	- 1.0581 (-23.40)
β_{FCA9}	- 1.1616 (-24.79)
γ_0	- 1.1420 (-34.40)
$\sigma_{0-\varepsilon}$	5.5540 (33.15) ^a
σ_{0-G}	3.5257 (17.32) ^a

a. *t*-statistics apply to the logged parameter estimates.

8. Kling (1997) simulates nonzero covariances, but restricts the variances to be constant across the two types of data.

The model predicts choices correctly 73% of the 5,038 choice occasions (647 individuals \times 8 experiments = 5,176; subtract 138 missing yields 5,038). The pseudo- R^2 is 0.453. The mean predicted probability of the preferred alternative from the stated preference experiment is 0.63, with a standard deviation of 0.22.

An alternative is infrequently chosen when its probability of being chosen is small, and frequently chosen when its probability is high. For example, when the predicted probability of selecting alternative A is less than 0.1, A is chosen in only 5% of the pairs; but when the predicted probability is greater than 0.9, A is chosen in almost all of the pairs, 96%.

The parameter estimates from the model can be used to predict the conditional probability of choosing Green Bay under the hypothetical conditions over the individual's other (real) choices. This is Equation 6. Multiplying this probability by the actual number of open-water days for the respondent produces a lower-bound estimate of the number of Green Bay days under hypothetical conditions.

5. Estimated responses to and willingness to pay for the elimination of FCAs

The model can be used to estimate how the probability of fishing Green Bay will change (and hence how the number of days fishing Green Bay will change, holding constant total fishing days) from either a change in catch times or FCA levels. For example, holding constant other site characteristics, the probability of going to Green Bay would increase from 0.40 to 0.46 if its FCAs were eliminated. At an existing FCA Level of four, doubling the catch rate for perch would only cause an increase from 0.40 to 0.42. The FCAs by species for Level 4 are “no more than one meal per month” for trout/salmon, walleye, and bass. There is no advisory for perch.

Denote individual i 's expected compensating variation for a fishing day (to all sites) for a change in the characteristics of Green Bay, $E(CV_i^F)$, and denote individual i 's compensating variation for a Green Bay fishing day for a change in the characteristics of Green Bay, CV_i^G . The estimated CV_i^G and $E(CV_i^F)$, along with estimates of the current number of fishing days and Green Bay fishing days can be used to obtain two lower-bound estimates of WTP for the elimination of FCAs for this target population.

For an improvement in Green Bay, CV_i^G is how much the angler would pay per Green Bay fishing day for the improvement, and CV_i^F is how much the angler would pay per fishing day (for fishing days at all sites). Note that for an improvement in Green Bay, $0 \leq CV_i^F \leq CV_i^G$, and for a deterioration in Green Bay, $CV_i^G \leq CV_i^F \leq 0$. An angler will pay no more per fishing day to have the FCAs at Green Bay reduced than he would pay per Green Bay fishing day because all fishing days are not necessarily to Green Bay.

For an improvement in Green Bay conditions, $CV_i^G \times D_i^{G^0} \leq CV_i^F \times D_i^{F^0} \leq CV_i^G$, where $D_i^{G^0}$ is the number of days in a season individual i fishes Green Bay under current (injured) conditions, and $D_i^{F^0}$ is the number of days individual i fishes all sites under current conditions (Morey, 1994).⁹

Since CV_i^G is per Green Bay fishing day conditional on fishing Green Bay, CV_i^G is not a random variable and can be estimated. The random components cancel out of the CV formula when the individual chooses the same alternative in each state. In discrete choice models

9. Given the model, CV^F and CV^G are constants independent of the individual's number of fishing days and Green Bay fishing days. This follows from the assumption that the utility from a fishing day (Green Bay fishing day) is not a function of the number of fishing days (Green Bay fishing days). In this case, any quality increase can be represented by an equivalent price decrease, and the inequality holds if the marginal utility of money is positive, which it is. That is, the inequality holds because the angler will not decrease fishing days if Green Bay improves in quality.

without income effects, the compensating variation can be written as the difference between the maximum utility in the two states multiplied by the inverse of the constant marginal utility of money:

$$CV_i^G = \frac{1}{\beta_y} (U_i^{G^1} - U_i^{G^0}) = \frac{1}{\beta_y} (\beta' x_i^{G^1} - \beta' x_i^{G^0}) \quad (12)$$

where $U_i^{G^1}$ is the utility from a Green Bay fishing day in the improved state, and $U_i^{G^0}$ is the utility in the current state; that is, G^1 denotes Green Bay under improved conditions and G^0 denotes Green Bay under current conditions. In addition, $x_i^G = x^G \forall i$, so $CV_i^G = CV^G \forall i$.

\hat{CV}^G (the estimated value of CV^G) for reducing FCAs from FCA Level 4 (the current level) to FCA Level 1 (no FCAs) is \$9.75 [\$8.06-\$11.73]; that is, \$9.75 for every Green Bay fishing day. For comparison, \$9.75 is 13% of the average reported cost of a Green Bay fishing day. \hat{CV}^G for reducing FCAs from Level 3 to Level 1 (no FCAs) is \$4.86. For reducing FCAs from FCA Level 2 to no FCAs, it is \$1.81. For comparison, \hat{CV}^G for doubling the perch catch rate is \$3.72, for quadrupling it is \$5.58, for a ten-fold increase it is \$6.97, and for doubling the catch rate of all species it is \$12.78.

Since CV_i^F is per fishing day and on each fishing day the angler has the choice of two sites: Green Bay or elsewhere, CV_i^F is a function of unobservable stochastic components, and so cannot be estimated. Instead we estimate its expectation:

$$E(CV_i^F) = \frac{1}{\beta_y} \left[E(\max(U_i^{G^1}, U_i^O)) - E(\max(U_i^{G^0}, U_i^O)) \right] \quad (13)$$

where U_i^O is the utility from fishing at another site. Given that U_i^G and U_i^O are bivariate normal:

$$E(\max(U_i^G, U_i^O)) = \gamma_0 + (\beta' x_i^G - \gamma_0) \Phi\left(\frac{\beta' x_i^G}{\sigma_{0-G}} - \frac{\gamma_0}{\sigma_{0-G}}\right) + \sigma_{0-G} \phi\left(\frac{\beta' x_i^G}{\sigma_{0-G}} - \frac{\gamma_0}{\sigma_{0-G}}\right) \quad (14)$$

where $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function, $\phi(\cdot)$ is the standard normal density function (Maddala, 1983, p. 370), and

$$\sigma_{0-G}^2 = \text{Var}[\varepsilon_{ij}^0 - \varepsilon_{ij}^G] = \sigma_0^2 + \sigma_G^2 - 2\sigma_{G0} \text{ (see Section 3).}$$

Substituting Equation 14 into 13, and simplifying it one obtains:

$$\begin{aligned} E(CV_i^F) &= \frac{1}{\beta_y} [(\beta' x_i^{G^1} - \gamma_0) \Phi\left(\frac{\beta' x_i^{G^1} - \gamma_0}{\sigma_{0-G}}\right) + \sigma_{0-G} \phi\left(\frac{\beta' x_i^{G^1} - \gamma_0}{\sigma_{0-G}}\right) \\ &\quad - (\beta' x_i^{G^0} - \gamma_0) \Phi\left(\frac{\beta' x_i^{G^0} - \gamma_0}{\sigma_{0-G}}\right) - \sigma_{0-G} \phi\left(\frac{\beta' x_i^{G^0} - \gamma_0}{\sigma_{0-G}}\right)] \end{aligned} \quad (15)$$

Since, in this model, $x_i^G = x^G \forall i$, $E(CV_i^F) = E(CV^F) \forall i$.

\hat{CV}^F for reducing FCAs from Level 4 (the current level) to Level 1 (no FCAs) is \$4.17 [\$3.41-\$5.00]; that is, \$4.17 for every fishing day. Remember that \$4.17 applies to all fishing days, not just Green Bay fishing days, so it is less than \hat{CV}^G , which is \$9.75.¹⁰ The 95% confidence interval on the \$4.17 estimate is \$3.41 to \$5.00. \hat{CV}^F for reducing FCAs from FCA Level 3 to no FCAs is \$2.15, and for reducing FCAs from FCA Level 2 to no FCAs is \$0.82. For comparison, \hat{CV}^F for doubling the perch catch rate is \$1.52, for quadrupling it is \$2.32, for a ten-fold increase it is \$2.80, and for doubling the catch of all species it is \$5.58.

For comparison purposes, we also estimated the model without the stated frequency data; that is, using only the RP use data and the SP choice data. The results are very similar and not significantly different. \hat{CV}^F for reducing FCAs from Level 4 (the current level) to Level 1 (no FCAs) is \$4.19 [\$3.34-\$5.20], as compared to \$4.17 from the full model. \hat{CV}^G for reducing FCAs from FCA Level 4 (the current level) to FCA Level 1 (no FCAs) is \$10.30 [\$8.15-\$12.78]

10. Both \$9.75 and \$4.17 fall within the range of values in the literature. See, for example, Herriges et al. (1999), Chen and Cosslett (1998), Jakus (1998), and Parsons et al. (1999).

as compared to \$9.75 from the full model. If one estimates a model with only the SP choice data one cannot estimate $C\hat{V}^F$; there is no frequency data. $C\hat{V}^G$ for reducing FCAs from FCA Level 4 to FCA Level 1 from this simple probit model is \$10.29.

6. Conclusions

To take advantage of the relative strengths of different types of data, SP choice data and SP frequency data are combined with RP data to estimate preferences for a unique fishing resource. Estimation is generalized by allowing for nonzero covariance between the various stochastic components associated with the different types of data. This is facilitated by specifying a probit rather than a logit or nested-logit framework, but one could adopt either of the latter two specifications. The unique combination of data not only allows for the estimation of how individuals would trade off different characteristics of a commodity that is unique, but also how they would be expected to change the relative quantities they consume when characteristics change.

References

Adamowicz, W.L., J. Louviere, and J. Swait. 1998. Introduction to Attribute-Based Stated Choice Methods. Final Report, Resource Valuation Branch, Damage Assessment Center, National Oceanic and Atmospheric Administration. Submitted by Advanis.

Adamowicz, W., J. Louviere, and M. Williams. 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 26:271-292.

Adamowicz, W.L., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management* 32:65-84.

Amemiya, T. 1994. *Introduction to Statistics and Econometrics*. Cambridge: Harvard University Press.

- Batsell, R.R. and J.J. Louviere. 1991. Experimental Analysis of Choice. *Marketing Letters* 2:199-214.
- Ben-Akiva, M. and T. Morikawa. 1990. Estimation of Travel Demand Models from Multiple Data Sources. In *Transportation and Traffic Theory*, M. Koshi (ed.). New York: Elsevier. pp. 461-76.
- Ben-Akiva, M., M. Bradley, T. Morikawa, J. Benjamin, T. Novak, H. Oppewal, and V. Rao. 1994. Combining Revealed and Stated Preference Data. *Marketing Letters* 5(4):335-350.
- Boxall, P. and W. Adamowicz. 2002. Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics* 23(4):421-446.
- Bradley, M.A. and A.J. Daly. 1991. Estimation of Logit Choice Models Using Mixed Stated Preference and Revealed Preference Information. Presented to the 6th *International Conference on Travel Behavior*, Quebec, May 22-24, 1991.
- Breffle, W.S. and R.D. Rowe. 2002. Comparing choice question formats for evaluating natural resource tradeoffs. *Land Economics* 78(2):298-314.
- Cameron, T.A. 1992. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Economics* 68(3):302-317.
- Cattin, P. and D.R. Wittink. 1982. Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing* 46:44-53.
- Chen, H.Z. and S.R. Cosslett. 1998. Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach. *American Journal of Agricultural Economics* 80:512-520.
- Englin, J., and T.A. Cameron. 1996. Augmenting Travel Cost Models with Contingent Behavior Data. *Journal of Agricultural and Resource Economics* 21(1): 130-149.
- Green, P.E. and V. Srinivasan. 1990. Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing* October:3-19.
- Hensher D.A. 1994. Stated Preference Analysis of Travel Choices: The State of Practice. *Transportation* 21:107-133.
- Hensher, D.A. and M. Bradley. 1993. Using Stated Choice Data to Enrich Revealed Preference Discrete Choice Models. *Marketing Letters* 4:139-152.
- Herriges, J.A., C.L. Kling, and D.J. Phaneuf. 1999. Corner Solution Models of Recreation Demand: A Comparison of Competing Frameworks. In *Valuing Recreation and the Environment*, J.A. Herriges and C.L. Kling (eds.). Northampton, MA: Elgar. pp. 163-197.

Jakus, P.M., D. Dadakas, and J.M. Fly. 1998. Fish Consumption Advisories: Incorporating Angler-Specific Knowledge, Habits, and Catch Rates in a Site Choice Model. *American Journal of Agricultural Economics* 80(5):1019-1024.

Kling, C.L. 1997. An Evaluation of the Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods. *Land Economics* 73:428-437.

Louviere, J.J. 1988. Conjoint Analysis Modeling of Stated Preferences. *Journal of Transport Economics and Policy* 10:93-119.

Louviere, J.J. 1992. Experimental Choice Analysis: Introduction and Overview. *Journal of Business Research* 24:89-95.

Louviere, J.J. 1996. Combining Revealed and Stated Preference Data: The Rescaling Revolution. Prepared for the AERE Workshop: Tahoe City, CA. June 2-4.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Mathews, K.E., W.H. Desvousges, F.R. Johnson, and M.C. Ruby. 1997. Using Economic Models to Inform Restoration Decisions: The Lavaca Bay, Texas Experience. TER technical report prepared for presentation at the Conference on Restoration of Lost Human Uses of the Environment, Washington, DC. May 7-8.

McFadden, D. 1986. The Choice Theory Approach to Market Research. *Marketing Science* 5(4):275-297.

Morey, E.R. 1994. What Is Consumer's Surplus Per Day of Use? When Is It a Constant Independent of the Number of Days of Use? and What Does It Tell Us about Consumer's Surplus. *Journal of Environmental Economics and Management* 26(3):257-270.

Morikawa T., M. Ben-Akiva, and D. McFadden. 1990. Incorporating Psychometric Data in Econometric Travel Demand Models. Prepared for the Banff Invitational Symposium on Consumer Decision Making and Choice Behavior.

Morikawa T., M. Ben-Akiva, and K. Yamada. 1991. Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data. *Transportation Research Record* 1328:30-35.

Parsons, G.R., P.M. Jakus, and T. Thomasi. 1999. A Comparison of Welfare Estimates from Four Models for Linking Seasonal Recreational Trips to Multinomial Logit Models of Site Choice. University of Delaware working paper (April).

Swait, J. and J. Louviere. 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* XXX (August): 305-314.

Swait, J., J. Louviere, and M. Williams. 1994. A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice. *Transportation* 21:135-152.

Wittink, D.R. and P. Cattin. 1989. Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing* 53:91-96.

Appendix: Derivation of the conditional probability

Consider the probability of choosing Green Bay site k_{ij} over a non-Green Bay site, conditional on the choice of Green Bay site k_{ij} over Green Bay site $3 - k_{ij}$. To ease the notation, suppose alternative 1 is chosen rather than alternative 2, and the individual and choice occasion subscripts are ignored. Under assumptions 2 and 3, the random vector $(\varepsilon^1, \varepsilon^2, \varepsilon^0)$ has a multinormal distribution with zero mean vector and covariance matrix:

$$\begin{pmatrix} \sigma_\varepsilon^2 & 0 & \sigma_{\varepsilon 0} \\ 0 & \sigma_\varepsilon^2 & \sigma_{\varepsilon 0} \\ \sigma_{\varepsilon 0} & \sigma_{\varepsilon 0} & \sigma_0^2 \end{pmatrix} \quad (A1)$$

This implies:

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \varepsilon^0 - \varepsilon^1 \\ \varepsilon^2 - \varepsilon^1 \end{pmatrix} \sim N(0, \Omega) \quad (A2)$$

where:

$$\Omega = \begin{pmatrix} \sigma_{0-\varepsilon}^2 & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & 2\sigma_\varepsilon^2 \end{pmatrix} \quad (A3)$$

The probability in Equation 6 is a conditional probability of a bivariate normal random variable, where the conditioning event does not have zero probability (which is the more usual case).¹¹ Let $a_1 = -\beta'(x_{ij}^0 - x_{ij}^{k_{ij}})$ and $a_2 = -\beta'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})$. From Amemiya (1994, pp. 35-36), denoting the joint, marginal, and conditional density functions of ω and its elements as f , we have:

11. It is a conditional probability, rather than a conditional expectation, so the Mill's ratio results from the selection literature (e.g., Maddala, 1983, p. 367) cannot be used.

$$f(\omega_1|\omega_2 < a_2) = \frac{\int_{-\infty}^{a_2} f(\omega_1, \omega_2) d\omega_2}{P(\omega_2 < a_2)} \quad (\text{A4})$$

so that:

$$P(\omega_1 < a_1 | \omega_2 < a_2) = \int_{-\infty}^{a_1} f(\omega_1 | \omega_2 < a_2) d\omega_1 = \frac{\int_{-\infty}^{a_1} \int_{-\infty}^{a_2} f(\omega_1, \omega_2) d\omega_2 d\omega_1}{\int_{-\infty}^{a_2} \int_{-\infty}^{a_2} f(\omega_1, \omega_2) d\omega_2 d\omega_1} \quad (\text{A5})$$

This is the ratio of a bivariate normal cumulative distribution function evaluated at α_1 and α_2 to a univariate normal cumulative distribution function evaluated at α_2 :

$$P(\omega_1 < a_1 | \omega_2 < a_2) = \frac{\Phi_2(a_1 / \sigma_{0-\varepsilon}, a_2 / \sqrt{2}\sigma_\varepsilon; \rho)}{\Phi(a_2 / \sqrt{2}\sigma_\varepsilon)}, \quad (\text{A6})$$

which is Equation 6.