Marine Resource Economics, Vol. 4, pp. 95–109, 1987
Printed in the UK.
0739-1360/87 $3.00 + .00

A Simulation Approach to Comparing Multiple Site Recreation Demand Models Using Chesapeake Bay Survey Data

CATHERINE L. KLING

University of California, Davis
Department of Agricultural Economics

Abstract *To value water quality improvements in the Chesapeake Bay or elsewhere, it is necessary to choose an appropriate model of consumer behavior. A number of different travel cost based recreation demand models have been employed to value changes in water quality or beach access. Among the possible models to choose from are the typical trip model, the pooled observations approach, a varying parameter model, and a logit model. Each approach makes different assumptions about the structure of individual preferences and the choice process underlying individual decisions.*

The purpose of this paper is to implement a methodology that can be used to suggest a model (or models) appropriate for valuing quality improvements in the Chesapeake Bay. To compare these approaches, a series of outdoor recreation user populations is constructed by choosing a utility function, its parameter values and an error distribution. This information is combined with the characteristics of individuals and recreation sites from a Chesapeake Bay recreation demand survey to solve the individual's maximization problem. Each of the models is estimated using these data, and the compensating variation of a quality change is calculated. Benefit estimates are compared with simulated welfare change to evaluate the models.

There is a large and growing literature concerning the theoretical and empirical aspects of valuing improvements in environmental amenities. A number of recreation demand models have been developed and used to value potential losses of unique natural resources as well as changes in the quality of outdoor recreation sites. Although extensive work has been done concerning the estimation of these approaches, relatively little has been done concerning their suitability or soundness as a basis for environmental policy. Of interest here is the applicability of these valuation techniques to recreational uses of the Chesapeake Bay.

Recreational uses of the Bay such as swimming, boating and fishing are an important component in Chesapeake Bay management decisions. Cleanup or other preservation activities will affect the desirability of the water for these uses. The use of recreation demand models to determine these values will benefit policy makers only if the estimates resulting from these models are reliable. The purpose of this paper is to examine the reliability of these models for the valuation of Chesapeake Bay recreation areas.

To accomplish this objective, a simulation experiment is conducted. Chesapeake Bay survey data is combined with a utility function to solve a set of hypothetical individuals' maximization problems. The solution to the maximization problems yields the simulated data. Since the utility function is known, willingness to pay for a hypothetical change in quality can be calculated. Standard recreation demand models are estimated using the simulated data and the resulting welfare estimates are compared to the simulated willingnesses to pay. Repeated trials are undertaken in order to assess these welfare estimates. Since data concerning users of the Chesapeake Bay and recreation sites along the Bay are employed in these simulations, the results are especially applicable to research and policies concerning the Bay.

The Estimation Approaches

A number of different multiple site recreation demand models have been estimated in the environmental economics literature. Typically, researchers employ one of these models to estimate the total benefits due to an improvement in environmental quality. Each of the models uses different estimation techniques and behavioral assumptions. The econometric models considered here can be grouped into three categories: single equation models, a varying parameter model, and a logit model.

1. Single Equation Demand Models

Perhaps the most intuitive approach to estimating welfare associated with quality changes for several sites is to pool all of the observations on number of trips, price and quality together and estimate a single demand equation. By pooling observations in this manner, it is possible to estimate the effect of quality on visits since quality will vary over the observations. Several different methods have been presented along these lines.

One variety of the single equation model is a typical trip model where a demand equation is estimated using the sum of all visits to all sites as the dependent variable and the travel cost and quality characteristics associated with the "typical" trip as the independent variables (Caulkins, Bishop and Bouwes 1982). The typical trip is generally defined to be the trip most often taken or the trip indicated by the individual as most preferred. The demand function estimated is

$$X_i = \alpha_0 + \alpha_1\overline{q}_i + \alpha_2\overline{P}_i + \alpha_3 y_i + e_i, \quad i = 1, \ldots, M, \tag{1}$$

where Greek letters correspond to estimated coefficients, X_i is the total number of trips individual i takes in a season to all sites, $\overline{q}_i$ is the water quality measure associated with the typical trip, $\overline{P}_i$ is the travel cost associated with this typical trip, y_i is income, e_i is a random term and M is the number of individuals in the sample.

Once this equation is estimated it is possible to determine Marshallian consumer surplus estimates associated with an increase in quality. The consumer surplus resulting from an improvement in quality is estimated as the difference between the total consumer surplus before and after the change. That is

$$CS_i = -.5(\hat{X}1_i^2/\alpha_2) + .5(\hat{X}0_i^2/\alpha_2), \quad i = 1, \ldots, M, \tag{2}$$

where $\hat{X}1_i$ is the predicted number of trips after the improvement in quality and $\hat{X}0_i$ is the predicted number of trips before the quality increase.

A second variety of the single equation approach is to consider visits to each site as separate observations in the estimation of a demand function (Reiling, Gibbs and Stoevner 1973; Binkley and Hanemann 1975; and Freeman 1979). In this approach, a pooled demand function is estimated where the dependent variable is the number of visits to the site and the independent variables are income and the corresponding travel costs and quality characteristics. That is

$$x_{ij} = \alpha_0 + \alpha_1 q_{ij} + \alpha_2 P_{ij} + \alpha_3 y_i + e_{ij}, \quad i = 1, \ldots, M, j = 1, \ldots, N, \tag{3}$$

where x_{ij} is the number of visits to site j, P_{ij} is the travel costs to site j, q_{ij} is the quality at site j, and N is the total number of sites. Marshallian welfare estimates are calculated in a similar manner as for the typical trip model.

An extension of this model can be estimated by including the prices of substitute sites, perhaps in the form of a stacked regression (Kling, Bockstael, and Strand 1985). However, it is still not possible to include substitute qualities since doing so eliminates the necessary quality variation.

2. *Systems of Demands and the Varying Parameter Model*

Another approach to multiple site recreation demand modelling is to consider a system of demand equations for the available sites. This approach is taken by Burt and Brewer (1971) and Cicchetti, Fisher and Smith (1976). This model generalizes a single site travel cost model to a system of demand equations. Since the travel costs associated with visiting a site are interpreted as the price of visiting that site, each individual faces a different price gradient depending on where he lives. The estimated coefficients of the demand equation are then used to predict the future usage rates or to construct benefit measures associated with changes in recreation sites.

The difficulty with this approach is that site characteristics cannot be incorporated as separate variables since their values do not vary over the observations in the demand functions. These characteristics can be incorporated into a system of demand equations by means of a varying parameter model (Freeman 1979; Vaughan and Russell 1982; Smith, Desvousges and McGivney 1983; and Smith and Desvousges 1985).

The varying parameter model is used by Vaughan and Russel to estimate the average value of a freshwater fishing day at fee-fishing sites. To accomplish this, they estimate a system of demand equations where the number of visits is specified only as a function of own price and income

$$\begin{array}{ccccccccc} x_{i1} & = & \alpha_1 & + & \beta_1 P_{i1} & + & \gamma_1 y_i & + & e_{i1} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ x_{iN} & = & \alpha_N & + & \beta_N P_{iN} & + & \gamma_N y & + & e_{iN}, \qquad i = 1, \ldots M. \end{array} \tag{4}$$

Next, the parameter values from these demand equations are assumed to depend on the characteristics of the sites

$$\begin{aligned} \alpha_j &= \delta_{10} + \sum_k \delta_{1k} q_{kj} \\ \beta_j &= \delta_{20} + \sum_k \delta_{2k} q_{kj} \\ \gamma_j &= \delta_{30} + \sum_k \delta_{3k} q_{kj}, \end{aligned} \tag{5}$$

where q_{kj} is in the amount of characteristic k at site j and the δ's are the second step regression coefficients. By substituting (5) to (4), Saxonhouse (1979) shows that estimation of the following equation is equivalent to estimating (4) and (5) separately (when there are two quality characteristics)

$$x_{ij} = \sigma_0 + \sigma_1 p_{ij} + \sigma_2 q_{1j} + \sigma_3 q_{2j} + \sigma_4 P_{ij} q_{1j} + \sigma_5 P_{ij} q_{2j} + \sigma_6 y_i q_{1j} + \sigma_7 y_i q_{2j} + \sigma_8 y_i + \epsilon_{ij}, \quad i = 1, \ldots, M, j = 1, \ldots, N,. \tag{6}$$

This equation is estimated using pooled data from all the sites visited. The quality and price variables all enter alone and multiplicatively.

Applications of the varying parameter model have specified the demand for a site as a function of its own price, income, socioeconomic variables and through the regression coefficients, own site characteristics. Through this specification, the number of visits to a site is a function of own price and own quality, but not of the other chosen prices and qualities.

3. *The Logit Model*

A number of authors have used multinominal logit models to estimate benefits associated with recreation goods (Binkley and Hanemann 1978; Feenberg and Mills 1980; Rowe, Morey, Ross and Shaw 1985; Bockstael, Hanemann, and Strand 1986; and Caulkins, Bishop and Bouwes 1986). The logit model assumes that the individual is faced with a choice among discrete, quality differentiated alternatives and that on any given day the individual chooses the alternative that maximizes his utility. The probability of choosing an alternative can then be expressed as a function of the characteristics.

In this model, the indirect utility function associated with the choice of alternative j is written with one quality characteristic as

$$V_{ij} = V(P_{ij}, q_j) + w_{ij} = a + bP_{ij} + cq_j + w_{ij} \tag{7}$$

where a, b and c are parameters, and w_{ij} is an error term. The individual will choose to visit site j only if site j provides the greatest utility, i.e., if $V_{ij} > V_{ik}$ for all k. This implies that the probability that the individual will visit site j is

$$\Pi_{ij} = \text{Prob}(w_{ij} - w_{ik} > V(p_{ik}, q_k) - V(p_{ij}, q_j)), \quad \text{all } j \neq k. \tag{8}$$

If the w_{ij} follow the Weibull distribution, the logit model results and this probability can be written

$$\Pi_{ij} = \exp(V(p_{ij}, q_j)) / \sum_k \exp(V(p_{ik}, q_k)). \tag{9}$$

The likelihood function for a sample of individuals is the product of the individual probability statements over all individuals.

When there is a quality or price change, the individual reallocates his visits among the available alternatives. Hanemann (1982) shows that the compensating and equivalent variation for the logit model can be written

$$cv_i = ev_i = (1/b)(\ln \sum_k \exp(V_0) - \ln \sum_k \exp(V_1)) \tag{10}$$

where V_0 and V_1 are the indirect utilities before and after the change respectively. This provides an estimate of willingness to pay per choice occasion. An estimate of total willingness to pay per season can be attained by multiplying cv_i by the total number of trips taken annually. For a quality improvement this is likely to be an underestimate of benefits since as water quality improves, individuals are likely to take more trips.

Bockstael, Hanemann and Strand (1986) apply a Generalized Extreme Value variation of the logit model. This distribution is assumed to allow some correlation among the errors. To allow total number of trips to vary when a welfare change is considered, they employ a Tobit model to estimate a separate participation equation.

One difficulty with logit models is that they have considered each choice occasion to be independent of all the others. The total number of trips in a season is determined as a result of individual decisions made over the course of the year. This means that when it comes to prediction or benefit estimation, the researcher is forced to estimate a new equation to capture the effect a welfare change will have on total participation. Rather than coming from one underlying utility maximization problem, these two decisions have been estimated independently.

The Construction of the Simulation Experiment

The importance of generating the data cannot be overestimated since any conclusions that are drawn concerning the usefulness of the recreation demand models are only as good as the data used to test their worth. In this regard it is important that the data used be "reasonable." That is, are the values of prices, income, and number of visits good approximations of observed data? Are the characteristics of individuals observed in sensible groupings? In order to assure that the data provide a good representation of reality, actual data have been used whenever possible in the creation of the simulation experiment. The data on income and travel costs (prices) come from a survey conducted by Research Triangle Institute for the University of Maryland.[1] Approximately 400 Chesapeake Bay users were surveyed on site in the summer of 1984. Individuals provided information about their pattern of beach use to 17 beaches, the activities they undertook while visiting the beaches, and extensive personal data. Information on water quality characteristics was also collected.

The six most popular beaches where water quality characteristics are also available (unfortunately, these measures could not be associated with all of the beaches) are chosen to be included in the simulated data. The use of six beaches simplifies the creation of the data set and the subsequent model estimations substantially, but still provides enough variation in alternatives to adequately rep-

[1] A complete description of the survey can be found in Bockstael, Hanemann, and Strand, 1986.

resent the range of choices facing the individual consumer. The use of the Chesapeake data in this manner guarantees that characteristics are observed in sensible combinations.

The next step in creating a recreation data set is to specify a utility function, its parameter values, and an error structure. Once this is achieved, the personal and site characteristics can be combined with the assumed preference structure to solve each individual's utility maximization problem. This provides the data to be used in estimating each of the recreation demand models. Additionally, since the utility function is known, a simulated measure of compensating variation can be calculated for a postulated change in water quality for all individuals.

The utility function chosen for this data set is one suggested by Bockstael, Hanemann and Strand (1986)

$$U = \sum_j \Psi_j \ln(x_j - \theta) + \ln z \tag{11}$$

where $\Psi_j = \beta_0 + \beta_1 S_j + \beta_2 F_j + e_j$, β_0, β_1, β_2 and θ are parameters and z is the numeraire. The Ψ_j is a quality index which is a linear combination of two water quality measures, S_j and F_j and a random error, e_j. The quality measures, S_j and F_j, stand for secchi depth and fecal coliform count respectively. Each individual has the same preferences in that they all have the same utility function and parameter vector. The utility function (11) is a variant of the linear expenditure system with the parameters dependent upon site quality. The parameter θ is often interpreted as the subsistence level of the good and is constrained to be positive. This interpretation is not necessary and is counterintuitive in the recreation case. The θ is chosen to be negative which implies that recreation goods are not necessary for an individual's survival.

The linear expenditure system (11) is chosen for the simulation because it suggests behavior which is consistent with observed behavior of Chesapeake beach goers. First, as long as the θ is negative, the solution to the maximization problem yields corner solutions. Most Chesapeake Bay users consume only one or a few of the many sites available to them so the use of a functional form which allows corner solutions is critical. There are many utility functions for which this is not the case. Second, as long as the θ is negative, the own-price elasticities implied by the linear expenditure system are elastic. This appears to be consistent with available estimates of these elasticities which range from about $-.6$ to -2.1 (Strand 1986). In addition, since the utility function is additive, the income elasticities are positive which is also consistent with expectations and available estimates of actual income elasticities.

The income and prices (travel costs) are from the Chesapeake data and mean income is about \$40,000.[2] The environmental data come from actual measures of water quality associated with the six sites. For the applications here, the measures are defined so that increasing numbers represent improvements in quality. Secchi depth and fecal coliform counts are objective measures of water quality, but since secchi depth is a measure of turbidity and fecal coliform levels are indicators of

[2] The average travel costs and their standard deviations to each of the six sites are: (1) \$10.59 (4.86); (2) \$11.65 (6.10); (3) \$11.03 (5.70); (4) \$16.26 (8.11); (5) \$15.37 (6.06); and (6) \$16.24 (6.90).

odor, they are likely to reflect characteristics perceivable by recreationists. The connection between objective measures of water quality and perceptions is not well understood. Fortunately, for the simulation experiment this difficulty can be ignored since utility can be assumed to depend on objective measures.

The final consideration is the specification of the error structure in (11). The error term in this construction is considered a result of measurement error or random preferences on the part of individuals. The error appears in the quality index so there is one error associated with each alternative per individual. These terms are randomly generated and distributed according to a rectangular distribution ranging from 0 to .0001. These endpoints are chosen so that the errors are significant in determining the size of the quality index Ψ_j, but do not overwhelm the effect of the two quality variables. This specification is chosen for its simplicity.

The solution to the consumers utility maximization problem is generated using the parameters $\beta_0 = -.0001$, $\beta_1 = .0002$, $\beta_2 = .0000004$ and $\theta = -1$. A number of different parameter values were tried in the construction of the simulated data. These parameters are chosen arbitrarily to yield solutions to the maximization problem that are consistent with observed Chesapeake Bay data. For a detailed discussion of the simulation model and solution algorithm see Kling (1986). The parameter values are combined with five different sets of randomly drawn errors from a uniform distribution to create five independent data sets. Each data set has about 250 observations. The resulting data sets are summarized in Table I.

The solutions for the x's are roughly consistent with observed data on recreation visits in terms of magnitude. The actual number of trips taken to these six sites as reported in the Chesapeake Bay beach survey range from 0 to 80 with a mean across all six sites of 1.82. The simulated number of trips taken to these sites range from 0 to 40 with a mean across all sites of 2.73. The simulated visits are more evenly distributed than the observed data.

Once the maximization problem is solved and the simulation data is generated, it is possible to calculate compensating variation for proposed changes in the prices or quality characteristics. A ten percent increase in the levels of the two water quality variables (S_j and F_j) is proposed and evaluated. Compensating variation (CV) can be defined for the quality change as

$$V(p,q^0,y) = V(p,q^1,y-CV) \tag{12}$$

where q^0 represents the old level of the quality index and q^1 represents the new level (with the increased S_j and F_j). The values for CV are also reported in Table 1. The mean compensating variation associated with a ten percent improvement in water quality is about $17.24. The measures range from zero to about $11.00 for each of the repetitions.

Results of the Simulations

To determine how appropriate the four models described above are for estimating the benefits of environmental improvements along the Chesapeake Bay, the models are estimated using the simulated data. Once the estimated welfare measures are calculated, they are compared to the simulated compensating variations, underlying the simulated data. Since the emphasis of this work is on welfare

Table 1
Mean Number of Trips to Sites 1 through 6 and Consumer Surplus Associated with a Ten Percent Improvement in Water Quality

	Mean Number of Trips to Sites 1–6						Compensating
Data Set	x1	x2	x3	x4	x5	x6	Variation
1	3.30 (3.59)[1]	3.29 (3.77)	3.18 (3.55)	2.19 (3.45)	2.75 (4.02)	1.46 (2.54)	17.26 (18.01)
2	3.51 (3.98)	3.29 (4.00)	3.35 (3.97)	2.05 (2.92)	2.76 (3.84)	1.41 (2.17)	17.13 (18.34)
3	3.52 (4.04)	3.62 (4.74)	3.30 (3.65)	2.15 (3.20)	2.80 (3.94)	1.43 (1.44)	17.36 (18.27)
4	3.47 (3.77)	3.56 (4.45)	3.23 (4.09)	2.04 (3.10)	2.77 (3.90)	1.34 (1.98)	17.18 (18.22)
5	3.43 (3.71)	3.31 (3.76)	3.14 (3.43)	2.13 (3.06)	2.81 (4.24)	1.38 (2.11)	17.26 (18.25)
Mean	3.44	3.41	3.24	2.11	2.78	1.40	17.24

[1] The standard deviations of x_1–x_6 and the compensating variation are reported in parentheses next to the means.

evaluation, individual coefficient estimates are not reported; rather, the overall fit of the models are discussed in general.

The models chosen for inclusion in the simulation are those which have received the most attention in the multiple site recreation demand literature. For each approach, a number of decisions concerning functional form, method of estimation, and welfare calculation had to be made. These choices are meant to mimic choices made by researchers in other applications of the models. One difference between the applications here and elsewhere is that the demand functions are estimated using a linear form rather than the often employed semi-log form. Since the semi-log is asymptotic to the price axis, it presumes that individuals always purchase a positive amount of the good which is inconsistent with the model and data presented here.

1. Typical Trip Model

The first model estimated is the typical trip approach. One equation is estimated where the dependent variable is the sum of all the trips taken to all the sites in a season and the independent variables are income and the levels of quality and price associated with the typical trip. This typical trip is described by the qualities and cost associated with the most often chosen alternative. The estimated demand function is

$$X_i = \alpha_0 + \alpha_1\overline{S}_i + \alpha_2\overline{F}_i + \alpha_3\overline{P}_i + \alpha_4 y_i + e_i \quad i = 1, \ldots, M, \tag{13}$$

where $X_i = \sum_j x_{ij}$, $\overline{P}_i$, $\overline{S}_i$ and $\overline{F}_i$ correspond to the price and qualities of the trip most often taken and y_i is income.

Heteroskedasticity is often present and corrected for in applications of travel cost models because visits per person are likely to have a higher variance at higher income levels and at higher values of other explanatory variables. A commonly employed test of heteroskedasticity is the Park-Glejser test (Pindyck and Rubinfeld 1981). In this test the log of the predicted variance form the OLS estimation is regressed against the log of the explanatory variables of interest. Evidence of heteroskedasticity was found in all five cases, so weighted least squares is used to estimate Equation (13).

The R^2 statistics for the five repetitions range from .76 to .77. The coefficients on $\overline{S}$, $\overline{P}$ and y are all significant at the 1% level for all repetitions. The coefficients on $\overline{F}$ are significant at the 1% level for two of the five repetitions. Table II reports the coefficient estimates and t-statistics of the repetition which generates the closest welfare estimate for each of the models.

Consumer surplus associated with a ten percent increase in the two quality variables of the typical trip is calculated as the difference between the Marshallian welfare triangle before and after the improvement. Results for the five data sets are presented in Table III. The average consumer surplus estimates per person per season range from \$11.46 to \$15.36. In all five cases the welfare estimates are below the simulated measures.

2. Pooled Observation Demand Functions

A second model estimated is a pooled observation demand function where the independent variable is the number of trips taken to each site for each individual.

Table 2
Model Coefficients and t-statistics for the Repetition which Produces the Closest Welfare Estimates of each Model

Typical Trip $R^2 = .75$

$$X_i = \underset{(-.97)^*}{-5.90} + \underset{(1.29)}{.0063F_i} + \underset{(2.45)}{4.72S_i} - \underset{(-13.84)}{1.20P_i} + \underset{(26.35)}{.00052y_i}.$$

Pooled—WLS $R^2 = .53$

$$x_{ij} = \underset{(.73)}{.43} + \underset{(4.12)}{.71S_j} + \underset{(3.70)}{.0018F_j} - \underset{(-27.70)}{.29P_{ij}} + \underset{(31.04)}{.00009y_i}.$$

Pooled—Tobit $R^2 = .67$

$$x_{ij} = \underset{(-.25)}{-.18} + \underset{(5.44)}{1.13S_j} + \underset{(3.43)}{.0021F_j} - \underset{(-29.71)}{.42P_{ij}} + \underset{(32.79)}{.00012y_i}.$$

Varying Parameter Model $R^2 = .68$

$$x_{ij} = \underset{(-1.93)}{-3.31} + \underset{(2.52)}{1.42S_j} + \underset{(3.56)}{.0055F_j} + \underset{(.22)}{.025P_{ij}} + \underset{(1.68)}{.00005y_i}$$
$$+ \underset{(2.55)}{7.0E - 7F_jy_i} + \underset{(1.04)}{.00001S_jy_i} - \underset{(-4.61)}{.0005F_jP_{ij}} - \underset{(-1.53)}{.054S_jP_{ij}}.$$

Logit $\chi^2 = -1719.81$ (3 degrees of freedom)

$$V_{ij} = \underset{(4.90)}{.3018S_j} + \underset{(4.50)}{.00075F_j} - \underset{(-3.35)}{.137P_{ij}}.$$

* t-statistics are given in parentheses below the coefficients.

Table 3
Welfare Estimates for the Typical Trip, Pooled Demand Function, Varying Parameter Model and Logit Model[1]

Data Set	Simulated Compensating Variation	Typical Trip Consumer Surplus	Pooled Observation Consumer Surplus: WLS	Pooled Observation Consumer Surplus: Tobit	Varying Parameter Consumer Surplus	Logit Compensating Variation
1	17.26	12.01	7.98	6.50	6.60	6.42
2	17.13	11.46	10.60	10.97	13.44	10.46
3	17.36	15.36	10.52	10.48	13.06	9.50
4	17.18	14.49	13.44	14.36	7.01	13.53
5	17.26	14.79	10.33	15.64	11.40	10.99
Mean	17.24	13.62	10.57	11.57	10.30	10.18

[1] The welfare estimates for the typical trip model, the pooled observations models, and the varying parameter model are Marshallian measures; the simulated measures and logit estimates are compensated measures.

Instead of one observation per person as in the above two approaches, there is one observation per person per site in this approach. The estimated demand function is

$$x_{ij} = \alpha_0 + \alpha_1 P_{ij} + \alpha_2 S_j + \alpha_3 F_j + \alpha_4 y_i + e_{ij}, \quad i = 1, \ldots, M, j = 1, \ldots, N, \tag{14}$$

where x_{ij} is the number of visits to site j, S_j and F_j are the quality measures associated with site j and P_{ij} is the travel cost to site j. Again, heteroskedasticity is corrected for in all five cases. WLS is performed on (14) for all observations where x_{ij} is positive. The R^2 statistics for the WLS version of the pooled model range from .52 to .54. The coefficients on S, F, P, and y are all significant at the 1% level for all 5 repetitions. Coefficients are reported for the repetition which yields the closest welfare estimate in Table 2.

The average estimated Marshallian consumer surplus for each repetition is reported in Table 3. In cases where the predicted number of trips is negative, the number of trips is set to zero.[3] This provides an estimate of welfare per person per site, so to determine average consumer surplus per person per season, it is necessary to multiple these welfare estimates by the number of sites. The estimates are somewhat lower than the typical trip with a mean of \$10.57.

Equation (14) is also estimated using a Tobit model. The Tobit procedure is used to account for the bias introduced by the large number of zero trips to the sites. The R^2 statistics for the Tobit model range from .66 to .68 and all of the coefficients are significant at the 1% level. The resulting consumer surplus estimates are also reported in Table III. The use of the Tobit improves the welfare estimates somewhat, but the difference is not very notable except for case 5.

3. *Varying Parameter Model*

A third model estimated is a version of the varying parameter model. Since there are only six sites in the data sets, the two step version of the model is not feasible to estimate since there are only as many observations in the second regression as there are sites. One equation is estimated using observations pooled over sites. The two quality characteristics enter into the demand function crossed with price and income as well as individually.

$$\begin{aligned} x_{ij} = {} & \alpha_0 + \alpha_1 S_j P_{ij} + \alpha_2 F_j P_{ij} + \alpha_3 P_{ij} + \alpha_4 S_j + \alpha_5 F_j + \alpha_6 y_i \\ & + \alpha_7 S_j y_i + \alpha_8 F_j y_i + e_{ij}, \quad i = 1, \ldots, M, j = 1, \ldots, N,. \end{aligned} \tag{15}$$

Equation (15) is estimated using the Tobit procedure and the demand for visits after a ten percent improvement in both of the quality variables is predicted. The Marshallian welfare triangle is computed using a slightly different formula since

[3] As noted by a reviewer, the use of actual trips instead of predicted trips would have avoided this anomaly. However, Bockstael, Hanemann, and Strand (1986) argue that the use of actual values implies that the error term in (14) arises from omitted variables, since the actual value is not affected by variable omission. In the simulation experiment, there are no omitted variables; rather, the error is presumed to arise from random preferences which is consistent with employing predicted trips to calculate welfare estimates.

the slope of the demand function changes when quality changes. The consumer surplus per site associated with a change in quality is

$$cs_{ij} = .5(x0_{ij}^2/(\alpha_1 S_j^0 + \alpha_2 F_j^0 + \alpha_3)) - .5(x1_{ij}^2/(\alpha_1 S_j^1 + \alpha_2 F_j^1 + \alpha_3)) \quad (16)$$

where S_j^0 and F_j^0 represent the quality values before the improvement and S_j^1 and F_j^1 are the quality levels after the change. In order to have a downward sloping demand function it is necessary for $(\alpha_1 S_j + \alpha_2 F_j + \alpha_3) < 0$. When S_j and F_j are increased both the slope and the intercept of the demand function will be affected. If the demand function has a negative slope, the slope change will tend to increase welfare, and if the price intercept increases welfare will unambiguously increase. However, it is possible and in some cases here observed, for the intercept to fall so that, depending on the values for S_j and F_j, consumer surplus estimates can be negative.

The R^2 statistics for the varying parameter model range from .67 to .69. There is some problem with insignificance of individual coefficients due, most likely, to collinearity associated with the cross product terms. For example, for the first repetition, three of the coefficients are not significant at the 1% level. Coefficient estimates for the model which yields the closest welfare estimate are presented in Table 2.

The consumer surplus from (16) represents an estimate of the benefits associated with an increase in quality per person per site. As for the pooled model, these estimates are multiplied by six to provide an estimate of the benefits per season. Welfare estimates for the varying parameter model are reported in Table 3. On average, the varying parameter underestimates the welfare change by about $7.00.

4. Logit Model

A logit model is estimated using price and the quality measures as explanatory variables. The indirect utility function estimated is

$$V_i = aS_i + bF_i + cP_i \quad (17)$$

where V_i is the indirect utility associated with a visit to site i. Using the expression for the probability of choosing site i (9) the log of the likelihood function can be constructed and estimated. For all repetitions, a Chi-square statistic is calculated; these statistics indicate that the model is significant at the 1% level using a likelihood ratio test with three degrees of freedom. In addition, the individual coefficients on S, F, and P are significant at the 1% level for all the repetitions with the exception of the coefficient on F in repetition 1. Table II contains coefficient estimates for the repetition which yields the closest welfare estimate.

To capture the effect of the quality change on the number of trips individuals take in a season, a separate participation equation is estimated. The total number of trips taken is assumed to depend on income and an inclusive value (Domencich and McFadden 1975). The inclusive value captures information about the set of alternatives available to the individual. A compensating variation estimate (10) is obtained which provides an estimate of the willingness to pay for the improved quality per trip. This is then multiplied by the predicted number of trips resulting from the participation equation. The results are summarized in Table 3. The welfare estimates range from \$6.42 to \$13.53. Although underestimates, these predictions are fairly close with a mean of \$10.18.

The superiority of the typical trip model is surprising given the apparent mis-specification of the model. One possible explanation is that there is mis-specification of all of the models in terms of functional form. The demand function implied by the linear expenditure system with quality introduced as in (11) is not linear in all of the regressors, nor is the indirect utility function. However, the demand functions estimated for the typical trip, the pooled model, and the varying parameter model are linear. In addition, the indirect utility function estimated in the logit model is also linear. Since all of the estimated models use functional forms which do not match the LES, the fact that the typical trip model performs best may be purely coincidental.

A second possible explanation of the superiority of the typical trip model may lie in the simulated data. In this data, many individuals visit only a few of the available sites. The typical trip model employs data only from the most often visited site. This model becomes more appropriate as the number of sites visited decreases; when only one site is visited, there is no mis-specification at all.

A second surprising result is the consistently negative bias in the consumer surplus estimates. These underestimates may be due to the unmatching functional forms. Although the consumer surplus estimates are biased downwards in all cases here, there is no reason to assume that this will hold in general. These results are specific to the utility function, parameter values, and estimated functions chosen for this simulation and should not be generalized to other situations without more evidence. In fact, in a simulation experiment employing the same utility function but using different parameter values and error distribution, Kling (1986) found some overestimates of consumer surplus.

Conclusions

The primary purpose of this work is to compare different modelling approaches that might be used to estimate welfare changes for Chesapeake Bay beach sites. A linear expenditure system utility function is combined with individual data on income and travel costs to generate five simulated data sets used in the simulation experiment. Four of the most commonly applied recreation demand models are estimated using the simulated data. Welfare estimates from the recreation demand models are compared with the simulated welfare changes for a ten percent improvement in environmental quality across all sites.

The fairly naive typical trip model provides the closest estimates although they are consistently too small. On average, the typical trip model underpredicts the simulated welfare by about \$3.60. The Tobit version of the pooled model provides the second best estimate of welfare generating an average error of about −\$5.60. The WLS version of the pooled model, the logit model, and the varying parameter model generate average welfare estimates ranging from \$10.18 to \$10.57. The worst average estimate (the logit) captures only about 60% of the simulated welfare, and the best average estimate (the typical trip) captures about 80% of the simulated welfare.

An interesting result is that each approach underestimates, on average, the simulated welfare. This consistent underestimate suggests that the welfare estimates provide a lower bound to the simulated welfare. This conclusion is not warranted in general, however, since this underestimate may be due to the mismatch in functional form between the estimated models and the utility function used to generate the simulated data.

A number of modifications to the simulation experiment presented here could be undertaken to improve the reliability of the results. For example, individuals could be made to vary in more dimensions than just income and price. In addition to consideration of the attributes of individuals and alternatives, issues of functional form, error structure, and variable omission can be examined with a simulation approach. The utility function in (11) is just one example of a utility function that could be employed in the simulation. A different functional form could be chosen to examine how the choice of utility function affects the simulation results. The choice of the error structure in (11) may also influence the estimated welfare measures.[4]

Simulation experiments, such as the one presented here, seem to be a tractable way to compare different modelling approaches, in particular, those models used to estimate benefits from recreational uses of a Chesapeake Bay. The results of this experiment indicate that four commonly employed recreation demand models underestimate the benefits of a ten percent improvement in environmental quality capturing from about 60 to 80% of the simulated welfare change.

Acknowledgments

This is Gianni Foundation Research Paper No. 849. The helpful comments of N. Bockstael, M. Copper, L. Deck, W. Oates and two anonymous referees are gratefully appreciated.

References

Binkley, C. S., and W. M. Hanemann 1978. *The recreation benefits of water quality improvements: an analysis of day trips in an urban setting*. Washington, D.C. EPA 60075-78-010.

Bockstael, N., M. Hanemann, and I. Strand 1986. *Measuring the benefits of water quality improvements using recreation demand models*, Report presented to the EPA under Cooperative Agreement CR-81143-01-1, Volume II.

Burt, O. and D. Brewer 1972. Estimation of net social benefits from outdoor recreation. *Econometrica* 39:813.

Caulkins, P., R. Bishop, and N. Bouwes 1986. The travel cost model for lake recreation: a comparison of two methods for incorporating site quality and substitution effects. *American Journal of Agricultural Economics* 68:291–97.

Cicchetti, C., A. Fisher, and V. K. Smith 1976. An econometric evaluation of a generalized consumer surplus measure: the mineral king controversy. *Econometrica* 44:1259–76.

[4] One way to introduce more variation and realism into the data is to let the parameters of the utility function vary over individuals. In addition, instead of using the rectangular distribution the normal distribution (or any other) might be employed. To examine the importance of these two variations, the simulation experiment is repeated allowing the parameters to vary over individuals and employing the normal distribution for the e_j's. The results are consistent with those presented here, i.e., the relative performance of the models is unchanged by these additions.

Domencich, T., and D. McFadden 1975. *Urban travel demand: a behavioral analysis.* North-Holland: Amsterdam.

Feenberg, D., and E. Mills 1980. *Measuring the benefits of water pollution abatement.* New York: Academic Press.

Freeman, M. A., III 1979. *The benefits of environmental improvement: theory and practice.* Baltimore: Johns Hopkins Press for Resources for the Future, Inc.

Hanemann, W. M. 1982. Applied welfare analysis with qualitative response models, California Agricultural Experiment Station working paper No. 241.

Harrison, D., and D. Rubinfeld 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5:81–102.

Just, R., D. Hueth, and A. Schmitz 1982. Applied welfare economics and public policy. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Kling, C. 1986. Measuring the recreational benefits of environmental amenities using multiple site models: an evaluation of methods. Ph.D. dissertation, University of Maryland.

Kling, C., N. Bockstael, and I. Strand 1985. An evaluation of multiple site model development. Paper presented at the AEA meetings, New York.

Maler, K. 1974. *Environmental economics: a theoretical inquiry.* Baltimore: Johns Hopkins University Press for Resources for the Future.

Pindyck, R., and D. Rubinfeld 1981. *Econometric models and economic forecasts*, New York: McGraw-Hill Book Company.

Reiling, S. D., K. C. Gibbs, and H. H. Stoevner 1973. *Economic benefits from an improvement in water quality.* Washington, D.C.: United States EPA.

Rowe, R., et al. 1985. *Valuing marine recreation fishing on the pacific coast.* Report prepared for National Marine Fisheries Service, NOAA.

Saxonhouse, G. R. 1977. Regressions from samples having different characteristics. *Review of Economics and Statistics* 59:234–37.

Smith, V. K., W. H. Desvousges, and M. P. McGivney 1985. The generalized travel cost model and water quality benefits: a reconsideration. *Southern Economic Journal* 371–81.

Smith, V. K., W. H. Desvousges, and M. P. McGivney 1983. Estimating water quality benefits: an econometric analysis. *Southern Economic Journal* 422–37.

Strand, I. 1986. Personal communication.

Vaughan, W. J., and C. S. Russell 1982. Valuing a fishing day: an application of a systematic varying parameter model. *Land Economics* 58:450–62.

Vaughan, W. J., and C. S. Russell 1982. *Fresh Water Recreational Fishing: The National Benefits of Water Pollution Control.* Washington, D.C.: RFF Inc.