INVESTIGACIONES ECONÓMICAS. vol. XXXII (1), 2008, 5-25

USAGE AND VALUATION OF NATURAL PARKS IN CATALONIA, 2001-2002

MICHAEL CREEL
MONTSERRAT FARELL
Universitat Autònoma de Barcelona

We present results on the use and valuation of parks and natural areas in Catalonia. A structural random utility model is used to estimate utility per recreational trip, and a reduced form model is used to estimate the number of trips. The two together give an estimate of the overall flow of utility over course of the season. Welfare effects due to changes in availabilities of sites can be deduced from changes in overall utility. For many of the sites, compensating variation for loss of access is estimated to be of the magnitude of dozens of millions of euros, annually.

Keywords: recreation; random utility model; parks.

(*JEL* Q26, C25, C51, D61)

1. Introduction

This paper presents results of a survey-based study of the use and valuation of parks and natural areas in Catalonia. We explore the factors which determine the overall level of usage of natural areas, as well as the determinants of the selection of which areas to visit. From the model that explains visitation and site selection, we are able to recover estimates of valuation (concretely, compensating variation) for the annual usage of the parks and natural areas.

The data was gathered in Spring of 2002, and contains information regarding visitation in the year previous to the point at which the survey was completed. Thus, the analysis of visitation and valuation is on an annual basis, and we expect that the flow of value over time will be reasonably close to the flow that is inferred from the annualized

We thank a co-editor and two anonymous reviewers for helpful comments, though any remaining errors are our own. This research was supported by grants SEJ 2006-00712/ECON and CONSOLIDER-INGENIO 2010 (CSD2006-00016).

estimates that we obtain, as long as conditions remain similar to those at the time the survey was administered.

The methodology used is a structural random utility model which allows analysis of site selection and expected utility flows per choice occasion, linked with a reduced form model of the frequency of choice occasions over the period of a year. The estimation of a plausible overall structural form model is not possible with the available data, since the dependency across choice occasions cannot be modeled. We argue below that this simplification is probably not of significance for the purposes of policy formulation.

Summarizing the findings, use value is found to be very substantial, reaching dozens of millions of euros annually for many of the sites. Use value is found to depend significantly upon age, education, and income, which suggests that distributional issues should be taken into account when formulating policy regarding protection of natural areas. We find that use value differs quite a bit between sites, in a way that conforms to prior expectations. One contribution of the study is that these differences are quantified.

In the following sections we describe how the data was collected, and the methodology that we use. Then we present the econometric models, and finally the results for visitation and valuation.

2. Data collection

2.1 Survey design and execution

The survey instrument is short, simple and is focused only on the information necessary to carry out the desired analysis. It can be administered in at most five minutes time. A short survey was used to encourage participation in the survey as well as to facilitate the respondents' maintaining their concentration, to maximize the quality of the data gathered. Respondents were limited to those 18 years of age or older, and they provided information about their own visits to parks and natural areas, as well as socio-economic information. The information collected by the survey is reported in the Appendix.

The survey was administered by telephone by Gama Investigaciones, S.A., during the month of May, 2002. Telephone contacts were stratified by district of residence (comarca) to achieve representativeness in terms of the geographic distribution of population in Catalonia, as well as to achieve an informative variation in travel distances to the destinations under consideration. Two thousand complete records were collected. In order to gather this information, a total of 6758 telephone calls were made. Of these, 2382 answered but refused to participate. In 281 cases there was no one 18 years or older available at the time. 1837 calls were to FAX or answering machines. In 258 cases the line was busy. The response rate among those contacted who were eligible to participate was $2000/4382 = 45.6\%$. Evidence reported below indicates that the sample seems to be quite representative of the general population.

2.2 Survey variables

The socio-economic information gathered is age, sex, education, family income and location of residence. Residence is recorded by administrative district (comarca). Respondents were asked about the number of visits they had made to 14 natural parks in Catalonia during the year preceding the survey (the period May 2001 - May 2002). The 14 destinations included in the survey comprise all of the National and Natural Parks of Catalonia at the time of the survey (Aigüestortes, Cadí-Moixerò, Zona Volcànica de la Garrotxa, Aiguamolls d'Empordà, La Delta de l'Ebre, la Muntanya de Montserrat, Cap de Creus, Isles Medes, Ports de Beseit, Sant Llorens i l'Obac, Montseny), as well as three parks with other protection status that are included due to their wide use amongst the population (Garraf, Montnegre and Collserola).[1] No doubt many respondents visited other less important natural areas in Catalonia, or parks outside of Catalonia, but it was necessary to limit the scope of the study in some way. We expect that this limitation will introduce some bias in the analysis, but hopefully it will not be too substantial due to the fact that the major parks have been included.

2.3 Constructed variables

Residence is recorded by administrative district[2], of which there are 41. Travel distance and time between the capital of the district and the 14 parks was determined using geographical information system (GIS)

[1] For a map that shows the location of these areas, see
http://mediambient.gencat.net/cat/elmedi/parcsdecatalunya/
[2] Specifically, by *comarca*. A map is available at
http://www.comarcalia.com/general/default.asp?xxi=5

route finding software that calculates the distance and travel time of the fastest route between two points. The largest of the districts has an area of approximately $1700km^2$ and most are considerably smaller. The radius of a circle of area $1700km^2$ is $23.2km$, which gives an idea of the potential for measurement error due to our method of calculating distance. There is a larger potential for relative measurement error in the distances that are calculated to be small.

3. Descriptive analysis

3.1 Socioeconomic variables

Table 1 presents basic descriptive statistics for these variables. The sample contains 58% women, which is several percentage points above the actual level in the population (about 51.2%). The average age is 44 years, average education corresponds to roughly 11 years (almost complete secondary education) and the average family income level is about 16500 euros annually. The distribution of these variables in the sample corresponds quite well with the population distribution, according to the statistical information available from the Institut d'Estadística de Catalunya at the page http://www.idescat.es/idescat.htm and publications listed there. Thus we conclude that the sample is quite representative in terms of these variables.

TABLE 1
Descriptive Statistics: Socioeconomic variables

	mean	st. dev	min	max
Age	44.443	15.882	18.000	85.000
Education	10.804	4.741	0.000	18.000
Income	16550.063	8136.281	2999.400	47990.402
Sex	0.577	0.494	0.000	1.000

3.2 Visitation to parks

Table 2 presents summary statistics for annual visits to the parks. The most visited parks are Collserola, Montserrat, and Montseny, which are all semi-mountainous parks that are fairly close to the Barcelona metropolitan area, which contains more than half the population of Catalonia. The least visited parks are Ports de Beseit, Sant Llorens and Montnegre. One of these is a rugged mountainous area that is far from population centers, and the other two are less spectacular semi-mountainous areas that are probably of mostly local interest.

Table 3 shows the frequencies of annual visits by counts to each of the parks. For most parks, about 90% of the population never visits. In the case of Montserrat, which has a special religious significance, about 24% of the sample visits at least once. Collserola is another special case. It is visited by about 18% of the sample at least one time, and it has a higher frequency of higher counts as well. This is probably due to its proximity to the city of Barcelona.

TABLE 2
Descriptive Statistics: Visits

	mean	st. dev	min	max
Aiguestortes	0.106	0.377	0.000	5.000
Cadi Moixero	0.096	0.405	0.000	8.000
Zona Volcánica	0.235	1.389	0.000	50.000
Aiguamolls	0.129	0.528	0.000	12.000
Delta Ebre	0.226	0.560	0.000	5.000
Montserrat	0.353	1.216	0.000	40.000
Cap Creus	0.151	0.791	0.000	25.000
Isles Medes	0.127	0.833	0.000	25.000
Ports Beseit	0.046	0.375	0.000	11.000
San Llorens	0.061	0.357	0.000	10.000
Montseny	0.317	1.658	0.000	40.000
Montnegre	0.064	0.774	0.000	30.000
Garraf	0.181	1.129	0.000	20.000
Collserola	0.530	2.817	0.000	50.000

TABLE 3
Frequencies of Visits: All destinations

	mean	st. dev	min	max			
Aiguestortes	0.909	0.083	0.004	0.003	0.001	0.001	0.000
Cadi Moixero	0.924	0.064	0.008	0.002	0.001	0.001	0.001
Zona Volcánica	0.864	0.103	0.019	0.006	0.003	0.001	0.004
Aiguamolls	0.910	0.067	0.015	0.004	0.003	0.000	0.001
Delta Ebre	0.819	0.153	0.018	0.005	0.002	0.002	0.000
Montserrat	0.756	0.196	0.028	0.013	0.003	0.001	0.003
Cap Creus	0.898	0.085	0.009	0.003	0.003	0.001	0.002
Isles Medes	0.915	0.071	0.009	0.003	0.001	0.000	0.001
Ports Beseit	0.968	0.026	0.003	0.001	0.000	0.001	0.001
San Llorens	0.951	0.043	0.004	0.001	0.001	0.000	0.001
Montseny	0.831	0.130	0.021	0.007	0.003	0.002	0.006
Montnegre	0.963	0.031	0.003	0.002	0.000	0.000	0.001
Garraf	0.899	0.081	0.012	0.003	0.002	0.001	0.003
Collserola	0.823	0.107	0.032	0.013	0.008	0.004	0.014

In Table 4 we find information about visits to any of the parks, taken as a group, which is reflective of the overall demand for outdoor recreation at natural parks. About 55% of the sample makes at least one visit to one of the natural parks under consideration during the year, and the mode is one visit. Almost 5% of the population makes more than 10 visits during the year.

TABLE 4
Frequencies of Visits, Any Destination

Visits	Frequency
0	0.448
1	0.140
2	0.116
3	0.079
4	0.047
5	0.043
6	0.029
7	0.017
8	0.013
9	0.013
10	0.007
>10	0.048

4. Demand and valuation

We are concerned with explaining visitation to 14 natural areas in Catalonia, and with estimating the economic value of access to these sites. We have data on visitation to each of the sites during the course of one year. One could frame the decision of which sites to visit, and how many times, as a utility maximization problem that would define an optimal calendar of visits to the sites over the course of the year. This would be a complex problem, for a number of reasons. There could be contingencies due to learning about sites as a result of previous visitation. There could be inter-temporal effects that would affect the relative attractiveness of sites, conditional on previous visits. For example, a person who visited Montseny the previous weekend might be less inclined to repeat a visit to Montseny, compared to a person who made no trip the previous weekend. There might be habit formation. The likelihood of making a trip at any point in time would probably depend upon how long it has been since the last visit. The model might be formulated more naturally as a continuous time model - otherwise it would be unclear how to define when choice occasions occur, and how many occur over the course of a year. A serious attempt to formulate such a model would entail many complexities.

One serious drawback with this approach is that the data needed to implement it would be very costly to gather. One would need diary-type data of the pattern of visitations throughout the year. We do not have this sort of data, and such data sets are very rare. A second reason for not pursuing this sort of model is that it is probably not really needed to answer many interesting policy-relevant questions. For policy formulation, issues such as intra-seasonal changes in the attractiveness of sites are of minor importance at best - what is important for policy issues such as the provision of sites and the investment in their quality levels is the overall use of the sites, and the long run valuation of access to sites. These questions can probably be answered satisfactorily using a considerably simpler model that abstracts from intra-seasonal effects. This is the approach we take here.

Following Creel and Loomis (1992) and Hausman *et al.* (1995), we specify a linked site selection/trip frequency model to explain recreation choices over the course of the season. Let T indicate the total number of trips that are made. The trip frequency part of the model explains total demand for trips, T. The site selection part of the model explains how the selected number of visits are allocated to particular sites. Before discussing the trip frequency and site selection parts of the model, let us discuss the overall problem. We suppose that the flow of utility due to recreational trips during the course of the year is additively separable over trip events, and that the contribution to overall utility due to each trip event does not depend in any way on the other trip events. If this is the case, the order in which the trips are made does not matter, and the utility received from a visit to a given site does not depend upon the destinations or number of previous or future trips. Conditional on the number of trips taken, $T = n$, total indirect utility over the season may be broken down into a baseline utility that is due to consumption of other goods, and which depends on income (M), and the sum of the contributions from each of the trips, t_i,$i = 1, 2, ..., n$.

$$V(M, t_1, t_2, ..., t_n | T = n) = V_0(M) + \sum_{i=1}^{n} V(M, t_i) \qquad [1]$$

Importantly, note that the contributions to overall utility from each trip $V(M, t_i)$ are assumed not to depend upon the overall number of trips, n, nor on the history of realized trips. The only factor that is important, as we explain below, is the destination of trip t_i. While we believe that the separability and independence of historical ordering of the utility function is probably not a strictly correct assumption, we also believe that it is unlikely to lead to important biases when the object of analysis is overall seasonal behavior and valuation. For intra-seasonal analysis it might distort the results in an important way, but we do not do this sort of analysis here. It is a necessary assumption in order to perform valuation analysis with the sort of data we have, which does not record the date upon which trips are made, nor the order in which they are made.

We now turn to an overview of the parts of the model. The trip frequency model is used to estimate to overall demand for trips, T. Implicitly, demand for trips is the result of a utility maximization problem where the number of trips is chosen based upon preferences for recreational trips in relation to other goods. In our application, the trip frequency model is not explicitly derived from a model of utility maximization. It is a reduced form model, analogous to an ordinary Marshallian demand function for a good. In estimation, we simply seek a model that accurately and parsimoniously predicts the number of trips taken, taking into account that the realized number of trips is necessarily a natural number.

Once the number of trips is determined, the individual must allocate them to the 14 sites. The contribution of the jth trip $(j = 1, 2, ..., n)$ to overall utility depends upon which site is visited, since sites have different characteristics. To estimate the expected contribution to utility from a trip to any site, we need to estimate the probability that each of the 14 sites is chosen on each trip occasion. From this model we can obtain an estimate of expected utility received per trip occasion. The trip frequency model leads to an estimated density function for the number of trip occasions. Combining the two parts of the model, overall expected seasonal utility can be obtained. Specifically, the overall expected utility from trips over the course of the season is obtained by multiplying expected utility per trip, obtained from the site selection model, by the expected overall level of trips which is the result of the trip frequency model. Expected utility per trip does not depend upon the number of realized trips, due to the maintained assumption of separability. Welfare effects due to changes in availabilities of sites can be deduced from changes in overall utility over the recreation season. Next, we present the trip frequency and site selection parts of the model in more detail.

4.1 Trip frequency

The trip frequency model only seeks to explain the total number of trips made during the season, conditional on observable explanatory variables. It is a reduced form model, not explicitly derived from a utility maximization model. Given our assumptions regarding the utility function, embodied in equation [1], structural modeling of trip frequency is not needed for the purpose of estimating welfare and usage effects, so we avoid the unnecessary complication which it would entail. To find the best model to fit the data, we estimated a number of statistical specifications for count data, including Poisson, negative binomial, semiparametric, semi-nonparametric, hurdle and mixture models[3]. The favored model according to the consistent Akaike information criterion was a mixture of two negative binomial Type-I densities, with separate constants and dispersion parameters for the two components, but all slope parameters constrained to be equal across the two latent component densities[4]. The negative binomial density for a count random variable Y may be written as

$$f_Y(y|\phi) = \frac{\Gamma(y+\psi)}{\Gamma(y+1)\Gamma(\psi)} \left(\frac{\psi}{\psi+\lambda}\right)^{\psi} \left(\frac{\lambda}{\psi+\lambda}\right)^{y} \qquad [2]$$

where $\phi = \{\lambda, \psi\}$, $\lambda > 0$ and $\psi > 0$. When $\psi = \lambda/\alpha$, where $\alpha > 0$, we have the negative binomial-I model (NB-I), such that $E(Y) = \lambda$ and $V(Y) = \lambda + \alpha\lambda$.

A two component mixture negative binomial (MNB) is

$$f_Y(y, \phi^1, \phi^2, \pi) = \pi f_Y(y, \phi^1) + (1-\pi)\, f_Y(y, \phi^2), \qquad [3]$$

where $\pi > 0$. The properties of the mixture density follow in a straightforward way from those of the components. In particular, the mean of Y is $E(Y) = \pi\lambda^1 + (1-\pi)\,\lambda^2$. To accommodate conditioning variables, we parameterize the mean of each of the component densities as

$$\lambda^j = \exp(\beta_1^j + \beta_2 AGE + \beta_3 EDUC + \beta_4 M + \beta_5 SEX), \qquad [4]$$

[3] See Creel and Farell, 2005 for a review of these statistical models.

[4] The CAIC values for the best particular specifications for models of given types are: Poisson 13718.1; negative binomial 7392.5; Poisson semiparametric 7372.4; hurdle negative binomial 7418.6; semi-nonparametric negative binomial 7350.6; constrained mixed negative binomial (the model used here) 7347.6. The results we report in the paper change in no important way if the semi-nonparametric negative binomial (the runner-up in terms of CAIC values) is used.

$j = 1, 2$ (recall that M is income). The constant term, β_1^j varies across the two components of the mixture, but the other β parameters are fixed across components. The dispersion parameters α^j vary across the two components[5].

4.2 Site selection

To model the site selection problem, we use the multinomial logit (MNL) model. The MNL model implies independence across sites of the random components of utility, an assumption which may not be plausible in all circumstances. To check this, we also estimated the more general nested MNL (NMNL) model (McFadden, 1984). A simple two-level version of the NMNL model groups sites according to some *a priori* information into K mutually exclusive sets $A_k, k = 1, 2, ..., K$. The probability of a visit to a site j that belongs the set of sites A_k is

$$\Pr(t_i = j) = \Pr(t_i = j | t_i \in A_k) \Pr(A_k) \quad [5]$$

where

$$\Pr(t_i = j | t_i \in A_k) = \frac{\exp\left(x_j' \beta_j / \rho_k\right)}{\sum_{j \in A_k} \exp\left(x_j' \beta_j / \rho_k\right)}, \quad [6]$$

$$\Pr(A_k) = \frac{\exp\left(\alpha_k + \rho_k I_k\right)}{\sum_K \exp\left(\alpha_k + \rho_k I_k\right)}, \quad [7]$$

and

$$I_k = \log\left[\sum_{j \in A_k} \exp(x_j' \beta_j / \rho_k)\right] \quad [8]$$

(see Cameron and Trivedi, 2005, pg. 510, for example). The α_k are group-specific constants, one of which is normalized to zero to identify the model. We explored several specifications of the NMNL model where sites were grouped into one of two categories ($K = 2$), according to criteria that could be expected to be related to correlated preferences across sites (coastal/interior, mountainous/non-mountainous, *etc*). Estimation was by full information maximum likelihood (FIML), and care was taken to find the global maximum of the likelihood func-

[5] Deb and Trivedi (1997) give further details for constrained and unconstrained mixture negative binomial models.

tion[6]. In all cases, the β_k estimates were virtually identical (to several decimal places) to those from the simple multinomial logit model (discussed below). The ρ_k parameters were indistinguishable from one. The log-likelihood values of the NMNL and MNL models were identical[7]. The MNL model, being more parsimonious, was favored by the consistent Akaike information criterion (CAIC). For practical purposes, the estimated probabilities for the sites are identical independent of whether the NMNL or MNL models are used. One could further investigate possible correlations of errors across sites using more computationally intensive models such as the multinomial probit model or the random parameters multinomial logit model (McFadden and Train, 2000; Parsons *et al.,* 1998), but we believe that this would not be productive in the present case.

The result that the MNL model was favored was somewhat surprising to us, and to the reviewers. One may wonder what factors could have resulted in the NMNL model being favored. The MNL model implies that the odds ratio between two sites is constant with respect to changes in the composition of the set of alternatives, while the NMNL model allows the odds ratio to change. In the present case, the model with a constant odd ratios is favored. The strong effect of travel cost on choice probabilities (see below) may be dominating any existing correlations across sites. Also, few of the sampled individuals made more than two visits total (see Table 4), so we simply have little information about preferences for combinations of sites. If we had more data for which multiple visits were made, it would be possible to allow for individual specific effects, which could help to capture correlations in preferences across sites.

We now explain in more detail the MNL specification that was used for the reported results. Since the variables used to explain site selection are site-specific constants, which reflect differences in the overall attractiveness of sites due to their special characteristics, price, and

[6] Henscher (1986) notes that FIML and sequential estimation can produce substantially different estimates, and that FIML is to be preferred. He also notes that the NMNL model is not globally concave, which motivates the need for care in the maximization of the log-likelihood.

[7] A formal test of the parameter restrictions that imply collapse of the NMNL model to the MNL model requires care due to the fact that, under the null hypothesis, some of the parameters are restricted to the boundary of the parameter space. In the present case, we know that a proper test would not reject the restriction to the MNL model, since the log-likelihood values are identical.

income, the model is directly presented in this specific form. This model is based upon the hypothesis that the contribution to overall utility of the ith trip, the $V(M, t_i)$ in equation [1], supposing that the trip is to the jth site[8] may be written as

$$V(M, t_i) = \alpha_j + \beta(M - p_j) + \varepsilon_{ij}, \quad [9]$$

where p_j is the price of a visit to the jth site[9], and M is family income. We assume that the ε_{ij} are independent and identically distributed (both over time and sites) extreme value random variables. The α_j are site-specific constants that capture the special qualitative characteristics of each of the sites.

Under these assumptions, the probability that a site is visited is constant across all visitations, and is of the form

$$\Pr(t_i = j) = \frac{\exp(\alpha_j - \beta p_j)}{\sum_{k=1}^{J} \exp(\alpha_k - \beta p_k)} \quad [10]$$

(McFadden, 1984). Note that income drops out of the site selection probabilities, since the probability function is homogeneous of degree zero. For the same reason, it is necessary to fix one of the site-specific constants. In Section 5 we explain in more detail how p_j is defined. Given this form for the choice probabilities, it is a simple matter to estimate using maximum likelihood.

4.3 Estimating valuation

In this paper we focus on total use value of the 14 destinations. We seek to estimate the compensating variation for the loss of access to each of the destinations, in turn. Compensating variation has the interpretation that it is the quantity of money that would compensate for the loss of access to a site - it is willingness to accept (WTA) payment for loss of access.

Recall that, given that a trip is taken, the utility of a visit to the jth site is given in equation and the probability of a visit to the jth site is given in equation . Given this, it is easy to see that the expected utility of a trip is

$$E\left[V(M, t_i)\right] = \beta M + I, \quad [11]$$

[8] Recall that the index of choice occasions is $i = 1, 2, ..., n$ and the index of sites is $j = 1, 2, ..., 14$.

[9] The price of a visit to a site depends upon the distance and time needed to make a trip from the location of residence to the site. The details are explained below.

where

$$I = \log\left[\sum_{j=1}^{J} \exp(\alpha_j - \beta p_j)\right]. \quad [12]$$

The term I is known as the "inclusive value" (McFadden, 1984). It is an index of the expected utility associated with the possibility of making a choice from a set of alternatives. Total expected utility for the season is the expected number of trips, conditional on income and other covariates x, times expected utility per trip:

$$E(u) = E(T|x, M)\left[\beta M + I\right]. \quad [13]$$

The expected number of trips $E(T|x, M)$ is obtained from the trip frequency model, which in our case is the mixed constrained negative binomial model that was discussed above.

Now, if we consider loss of access to site j, the expected utility of a trip becomes

$$E\left[V(M, t_i)\right] = \beta M + I^*, \quad [14]$$

where

$$I^* = \log\left[\sum_{k=1, k\neq j}^{J} \exp(\alpha_k - \beta p_k)\right]. \quad [15]$$

Then total expected utility over the season after loss of access to the site is

$$E(u^*) = E(T|x, M)\left[\beta M + I^*\right]. \quad [16]$$

Expected compensating variation for loss of access to the site (CV) is the amount of money that will make expected utility after the loss of access to the site equal to initial expected utility with access to all sites. It is implicitly defined by

$$E\left[T\left(x, M + CV\right)\right]\left[\beta\left(M + CV\right) + I^*\right] = E\left[T\left(x, M\right)\right]\left[\beta M + I\right], \quad [17]$$

where the notation explicitly accounts for the possibility that expected trips may depend upon income, among other factors. A simple means of finding CV is to minimize the square of the difference between the two sides of the preceding equation. The minimum will be zero, and the minimizer is CV. A better intuition may be obtained by noting that if we ignore the income effect, which in fact is of minor importance, then $E\left[T\left(x, M + CV\right)\right]$ can be taken to be approximately equal to $E\left[T\left(x, M\right)\right]$ and equation simplifies to the approximation

$$CV \approx \frac{I - I^*}{\beta}. \qquad [18]$$

The change in expected utility of a trip due to loss of access to the site is translated into monetary terms by dividing by the (constant) marginal utility of income, β (see equation). This expression makes clear the dependence of CV on the value of β, to which we will return later.

5. Results

In this section we report the estimation results for the trip frequency and site selection models, and the valuation results for access to each of the sites.

Table 5 reports the estimation results for the CMNB-I trip frequency model, which was selected according to the CAIC criterion. Age, education, and income all have a positive and significant effect on the number of trips taken. The dummy variable for sex does not have a significant effect. The two constants of the two components of the mixture differ significantly from each other, which may be indicative of the existence of a sub-populations of visitors and non-visitors. The two over-dispersion parameters of the mixture components are both significantly different from zero and from each other. The estimated mixture probability is about 0.44. It is not estimated with very good precision. Note that a t-test that this parameter is equal to zero is not a valid test for the lack of a mixture, since the value of zero is on the boundary of the parameter space and the separate constants and overdispersion parameters are unidentified under the null. The important differences in the estimated values of the separate constants and over-dispersion parameters, as well as the fact that this model is favored by the CAIC, is evidence of the existence of two components.

To estimate the site selection model, the travel costs to the sites (the p_j of equations and) must be calculated. We define travel cost to a site as round trip distance times 0.40 euros per kilometer, plus 4.50 euros per hour of travel time, where travel time and distance were obtained using the GPS data mentioned above. This definition is admittedly somewhat arbitrary, and the sensitivity of the results to the definition warrants discussion. If this measure of travel cost is scaled by multiplying by a constant a, the estimated value of β will be divided by the same value, and the estimated value of compensating variation will in turn be multiplied by approximately a (see equation). A change in the relative weights of monetary cost and the opportunity cost of time will have a more complicated effect on the estimated valuation.

TABLE 5
Estimation Results: CMNB-I Trip frequency model

	estimate	st. err.	t. stat.	p-value
CONSTANT	-0.053	0.210	-0.252	0.801
AGE	0.009	0.002	3.936	0.000
EDUCATION	0.040	0.008	4.816	0.000
INCOME	0.000	0.000	2.656	0.008
SEX	0.020	0.065	0.303	0.762
ALPHA	22.497	0.278	80.971	0.000
CONSTANT2	0.904	0.066	13.697	0.000
ALPHA2	2.034	0.271	7.515	0.000
MIX	0.436	0.354	1.234	0.217

The estimation results for the site selection model appear in Table 5. The site-specific constant for Aigüestortes has been set to zero, to achieve identification, and the other site-specific constants are to be interpreted as reflecting attractiveness relative to Aigüestortes, apart from the effect of travel cost (which is subtracted from income). The marginal utility of income, the coefficient β in equation , is positive, as it should be, and strongly significant. Recall that there are 1838 observations. The value of the average log-likelihood function is -10.94, and the average CAIC value is 22.01. Most of the site specific constants are negative, which reflects the fact that the base site, Aigüestortes, the coefficient of which is normalized to zero, is a major National Park that is especially attractive.

TABLE 6
Estimation Results: MNL Site selection model

	estimate	st. err.	t. stat.	p-value
CADI MOIXERO	-1.046	0.121	-8.609	0.000
Z VOLCANICA	-0.294	0.165	-1.778	0.075
AIGUAMOLLS	-0.768	0.126	-6.118	0.000
DELTRA EBRE	0.185	0.109	1.704	0.088
MONTSERRAT	-0.546	0.129	-4.232	0.000
CAP CREUS	-0.311	0.147	-2.118	0.034
ISLES MEDES	-0.872	0.185	-4.724	0.000
PORTS DE BESEIT	-1.331	0.212	-6.277	0.000
SANT LLORENS	-1.957	0.159	-12.274	0.000
MONTSENY	-0.208	0.150	-1.386	0166
MONTNEGRE	-2.208	0.294	-7.509	0.000
GARRAF	-1704	0.169	-6.363	0.000
COLLSEROLA	0.364	0.132	-2.765	0.006
INCOME	0.013	0.001	20.249	0.000

5.1 Valuation

Next, we report results on valuation using the estimation results presented above. In all cases we report CV for the loss of access to a single one of the 14 sites. Note that loss of access does not have the same welfare implications as destruction of the site, since there may be a substantial non-use value component of overall value. In this report we seek to estimate only the use value component since the available data are not useful for estimation of non-use value. Using the estimated model, it would be possible to investigate the use valuation effects of more complicated forms of changes in conditions, such as loss of access to multiple sites, or the effect of the creation of a new site in a given location, but this has not yet been done, in part since so many different forms of changes may be imagined.

The specific way in which CV has been calculated has been to substitute the estimated coefficients of the site selection model and the CMNB-I trip frequency model into equation , which is then solved to find compensating variation. Note that in that equation, I^* changes depending upon which of the sites is eliminated. Also note that CV depends upon the values of the conditioning variables AGE, INCOME, EDUCATION, SEX, and upon the location of the residence of each individual, which determines the travel cost to each of the sites. Thus, we estimate a 1838×14 matrix of compensating variations, for each of the 1838 individuals and for each of the 14 sites, and these estimated compensating variations depend upon individual characteristics. Table 6 reports summary statistics for CV for the loss of access to the sites.

TABLE 7
Estimated compensating variation per capita: Summary statistics

	estimate	st. err.	t. stat.	p-value
AIGUESTORTES	3.936	8.867	0.832	93.771
CADI MOIXERO	2.544	1.182	0.656	10.533
Z VOLCANICA	6.142	3.089	1.093	22.791
AIGUAMOLLS	3.226	1.912	0.601	11.675
DELTRA EBRE	7.346	7.422	0.782	59.672
MONTSERRAT	9.642	2.935	2.054	18.172
CAP CREUS	3.776	2.229	0.729	15.385
ISLES MEDES	3.164	1.919	0.568	12.582
PORTS DE BESEIT	1.422	1.410	0.145	10.190
SANT LLORENS	1.569	0.421	0.439	3.703
MONTSENY	8.428	3.130	1.524	21.665
MONTNEGRE	1.597	0.502	0.186	3.285
GARRAF	4.828	1.772	0.591	10.934
COLLSEROLA	14.841	6.253	1.696	23.060

In Table 7 we see that, on average, the CV measure of use value per person per year is not very large, with values between about 1.40 to 14.80 Euros per person per year. Nevertheless, taking into account that the 18 year and older population of Catalonia in 2005 was approximately 5.812 million people (source: Institut d'Estadística de Catalunya, http://www.idescat.net/), the estimated annual values for the entire population of Catalonia would be very substantial, ranging from a low of approximately 8 million euros in the case of Ports de Beseit to 86 million euros in the case of Collserola.[10]. We also note that these estimates represent an annual flow of value, and that non-use value is not estimated here.

The sites with higher average valuation are the well-known destinations that are close to the Barcelona metropolitan area (Collserola, Montserrat, and Montseny, for example). Destinations that are more peripheral to the population center have lower valuation in general (for example Aigüestortes, Cadí-Moixerò or Ports de Beseit). Some destinations are characterized by relatively low valuation, with little variation, such as Montnegre and Sant Llorens de Munt, which are probably of mostly local interest since they have few unique qualities in relation to the better-known sites. Other sites such as Aigüestortes or the Delta de l'Ebre have substantial variation in valuations. The site specific constants that capture the effects of unique characteristics show that these two sites are quite attractive, apart from the effect of distance (recall that the site specific constants in Table 7 are relative to the zero value assigned to Aigüestortes). However, both sites are relatively far from the Barcelona metropolitan area, and this leads to a low valuation on the part of consumers from this area.

To obtain an idea of how CV depends upon other variables, we fit the response surface (Hendry, 1984) model

$$CV_{ij} = \beta_j + \beta_A AGE_i + \beta_E EDUCATION_i + \beta_I INCOME_i + \\ + \beta_S SEX_i + \beta_D DISTANCE + \varepsilon_{ij} \qquad [19]$$

by ordinary least squares. The results are in Table . The R^2 for this regression is 0.68. The AGE, EDUCATION, INCOME and SEX variables all have their sample means subtracted, so the site specific constants give an estimate of the sample mean CV when DISTANCE is zero. Here, it is clear that CV increases with AGE, EDUCATION,

[10]This assumes that the sample is representative of the population of Catalonia that is of age 18 or more. The use valuation of the population under 18 years of age is not estimated here.

INCOME and SEX, and declines in DISTANCE, when we control for the variables jointly. However, the magnitudes of the effects, except for DISTANCE[11], are quite small. The most important distributional result to note is that residential areas that are far away from any attractive recreational site are significantly (negatively) impacted.

Looking at the sites, Collserola, Montseny, Aigüestortes and Delta de l'Ebre are seen to be much appreciated, while sites such as Sant Llorens, Cadí-Moixerò and Garraf are less valued. When comparing Tables and it is important to keep in mind that the site specific constants in Table give an estimate of CV at the sample means of age, education and income, when distance (and thus, travel cost) is zero, while Table gives the estimated CV values unconditional on other variables. Thus, the latter Table averages CV values over all distances in the sample.

TABLE 8
CV as a Function of Other Variables

	estimate	st. err.	t. stat.	p-value
AIGUESTORTES	15.411	0.113	136.475	0.000
CADI MOIXERO	8.981	0.086	104.069	0.000
Z VOLCANICA	12.662	0.087	145.883	0.000
AIGUAMOLLS	10.582	0.091	116.787	0.000
DELTRA EBRE	16.171	0.098	165.339	0.000
MONTSERRAT	13.230	0.076	173.024	0.000
CAP CREUS	12.033	0.096	125.892	0.000
ISLES MEDES	9.993	0.092	109.107	0.000
PORTS DE BESEIT	10.399	0.102	102.381	0.000
SANT LLORENS	6.294	0.084	75.124	0.000
MONTSENY	13.899	0.083	167.437	0.000
MONTNEGRE	5.674	0.084	67.914	0.000
GARRAF	9.004	0.084	107.202	0.000
COLLSEROLA	17.835	0.081	220.358	0.000
AGE	0.009	0.001	6.556	0.000
EDUCATION	0.011	0.005	1.961	0.050
INCOME	0.015	0.003	5.038	0.000
SEX	0.183	0.054	3.383	0.001
DISTANCE	-0.023	0.000	-132.609	0.000

[11] The coefficient of DISTANCE gives the decline in CV per kilometer of round trip distance to the site.

6. Conclusions

This paper has presented an analysis of data on visitation to parks and natural areas in Catalonia during 2001/2002. We have analyzed the factors that explain visitation, and from the model for visitation we have inferred valuation of the 14 areas included in the study.

The model for visitation is formed of two parts - a trip frequency model, and a site selection model. We find that trip frequency is increasing in age, education and income. In the site selection model, site specific constants that capture the special characteristics of the 14 sites are important, as is income (and thus the travel cost of visiting the site).

When we move to an analysis of valuation, we see that the loss of access to a given site depends importantly upon which is the site that is closed to access. This is to be expected, both simply from common sense, and since site characteristics were found to be important in the site selection model. We find that compensating variation (willingness to accept payment in compensation for loss of access to a site) is increasing in age, education, and income, as declines in travel cost. Women have a slightly higher CV than do men.

We have seen that certain sites are valued considerably more highly than are others, which is not at all surprising. The estimated values essentially serve to quantify what was already probably known on a qualitative level. Perhaps the most important conclusion is that the access to natural parks and areas represents a very substantial contribution to social welfare. Estimated compensating variation for loss of access to the sites under study ranges from 8 million to 86 million euros annually. It is important to recall that the valuation of the under 18 years of age part of the population is not estimated here. Thus, the values estimated here, though they are substantial, are very likely downwardly biased estimates of the use benefits that the recreation sites provide to society.

Another important qualification is that the valuation results reported here are for *access* to the sites. This is only a part of the total valuation of the sites, which may depend in part upon the mere *existence* of the sites. It is perfectly plausible that people may attach a value to the existence of a site even if they never visit it. They may enjoy viewing photographs of the site, or they may value the fact that the site is preserved so their children may visit it in the future, for example. As such, the figures reported here should be taken as estimates of only a component of the total values of the sites.

The model we have worked with in this paper is necessarily somewhat limited, due to the limited information that the data gives us regarding the history of visits over the season. While it is fairly easy to imagine extensions to the model to allow for intra-seasonal effects, the required data would be considerably more costly to gather. However, some extensions to the present model that could be implemented with the available data are possible. One of the most interesting would be finding some means of more closely coupling the site selection and trip frequency components of the model. One could contemplate that trip frequency could depend upon the expected utility of a trip, which is closely related to the inclusive value that comes from the site selection component. Once the two components share parameters, efficient estimation would require joint estimation of the two parts. Our exploratory work on models of this sort was not very satisfactory, but we believe that this sort of extension could be of interest for future work.

Appendix. The survey instrument

This appendix gives the complete contents of the survey instrument, in abbreviated form. The original survey instrument is available from the authors upon request.

1. Asks age, and requests that telephone be passed to someone 18 or order if the respondent is a minor.

2. Asks number of visits to each of 14 destinations during the last year.

3. Asks a hypothetical question about visits to the Zona Volcanica de la Garrotxa (not used in this paper).

4. Asks a second hypothetical question about visits to the Zona Volcanica de la Garrotxa (not used in this paper).

5. Asks educational level of respondent (8 categories).

6. Asks annual family income (15 categories).

7. Asks “comarca” of residence (41 categories).

8. Notes gender of respondent (observed, not asked).

References

Cameron, A.C. and P.K. Trivedi (2005), *Microeconometrics,* Cambridge Univ. Press, New York.

Creel, M. and M. Farell (2005): "Likelihood-based approaches to modeling demand for medical care", UFAE and IAE Working Papers, number 646.05, http://econpapers.repec.org/paper/aubautbar/646.05.htm.

Creel, M. and J. Loomis (1992): "Recreation value of water to Wetlands in the San Joaquin Valley: Linked multinomial logit and count data trip frequency models", *Water Resources Research* 28, pp. 2597-606.

Deb, P. and P.K. Trivedi (1997): "Demand for medical care by the elderly: a finite mixture approach", *Journal of Applied Econometrics* 12, pp. 313-36.

Hausman, J.A., G.K. Leonard and D.L. McFadden (1995): "A utility-consistent, combined discrete choice and count data model: Assessing recreational use losses due to natural resource damage", *Journal of Public Economics* 56, pp. 1-30.

Hendry, D.F. (1984): "Monte Carlo Experimentation in Econometrics, in Handbook of Econometrics", Vol. II, Griliches, Z. and M.D. Intriligator (eds.), North Holland, Amsterdam, pp. 939-76.

Hensher, D. A. (1986): "Sequential and full information maximum likelihood estimation of a nested logit model", *The Review of Economics and Statistics* 68, pp. 657-667.

McFadden, D., (1984): "Econometric analysis of qualitative response models, in handbook of econometrics", Vol. II, Griliches, Z. and M.D. Intriligator (eds.), North Holland, Amsterdam, pp. 1396-1457.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response", *Journal of Applied Econometrics* 15, pp. 447-70.

Parsons, G. R., P. Jakus, and T. Tomasi (1999): "A comparison of welfare estimates from four models for linking seasonal recreational trips to multinomial models of site choice", *Journal of Environmental Economics and Management* 38, pp. 143-157.

Resumen

Este trabajo presenta los resultados obtenidos en la investigación realizada sobre el uso y valoración de los parques y espacios naturales en Cataluña. Se utiliza un modelo estocástico para estimar la utilidad por cada viaje de recreo. Por otra parte, se emplea un modelo reducido para estimar el número de viajes. Ambos proporcionan una valoración del flujo total durante la temporada. Los efectos sobre el bienestar producidos por cambios en la disponibilidad de distintos lugares, se pueden deducir a partir de los cambios en la utilidad global. Para muchas de las zonas, la variación compensatoria por la pérdida de acceso al lugar, se considera que asciende a docenas de millones de euros anuales.

Palabras clave: esparcimiento, modelos de utilidad estocástica, parques naturales.

Recepción del original, enero de 2006
Versión final, julio de 2007