Chapter 19

Estimating Forest Recreation Demand Using Count Data Models

Jeffrey E. Englin, Thomas P. Holmes, and Erin 0. Sills
University of Nevada - Reno, USDA Forest Service and North Carolina State University

Forests, along with related natural areas such as mountains, lakes, and rivers, provide opportunities for a wide variety of recreational activities. Although the recreational services supplied by forested areas produce value for the consumers of those services, the measurement of recreational value is complicated by the fact that access to most natural areas is non-priced. Because outdoor recreation often competes with commodity uses of forests, such as timber harvesting or mineral extraction, failure to account for the recreational use of forest land makes it impossible to determine the efficient use of forest resources.

A key insight attributed to Harold Hotelling is that the price of recreational access can be inferred from information on travel costs. Subsequent development of this idea was undertaken by Marion Clawson (1959) and, a few years later, articulated in a general work on the economics of outdoor recreation (Clawson and Knetsch 1966). The basic Hotelling-Clawson-Knetsch (HCK) approach to estimating recreation demand is to statistically regress the number of trips taken to a recreational site on the round-trip cost of travel between trip origins and the site. A set of demand shift variables are also typically included in the specification to control for socio-economic characteristics of visitors, indicators of site quality, and costs associated with visiting substitute sites. Once a travel cost demand curve is estimated, the value of a recreational site can be computed by integrating the area under the demand curve.

Two types of data can be used to estimate travel cost models (see, for example, Bockstael et al. 1991 and Freeman 1993). The early studies

Sills and Abt (eds.), Forests in a Market Economy, 341–359.

typically used aggregate data on origin zones; these are often referred to as zonal travel cost models. Per capita visitation rates for each origin zone (often counties, but also distance zones) were computed, and distances were translated into travel costs using cost per mile multipliers. Socio-economic variables for origin zones were proxy variables for the representative visitor, and prices based on travel costs to substitute sites were included in the specification.

The second type of data that can be used to estimate travel cost models is based on individual observations of visitation rates and socio-economic variables (referred to as individual travel cost models). Individual data do not rely on the representative visitor assumption. The added precision in describing individual characteristics and trip decisions has led to the development of a rich array of empirical methods and, in particular, models based on random utility maximization (RUM).

The RUM approach models the choice of a recreation site from among a set of alternative sites as a utility-maximizing decision, where utility includes a stochastic component. RUM models emphasize the impact of site quality on recreation demand and are estimated using either multinomial or nested logit models. Forestry examples include Englin et al. (1996) and Pendleton and Shonkwiler (200 1).

Another approach that focuses attention on site quality is the hedonic travel cost (HTC) method. The HTC method is used to estimate the demand for site characteristics using a two-step procedure (Brown and Mendelsohn 1984). In the first stage, marginal values (implicit prices) of the site characteristics are estimated for each origin zone. Then, demand functions for characteristics are estimated in the second stage across all origins. Applications to forestry include Englin and Mendelsohn (1991), Holmes et al. (1997), and Pendleton et al. (1998).

During the past decade, there has been an explosion of interest in the application of count data models based on the Poisson distribution to estimation of HCK models of recreation demand. In this chapter, we provide an overview of the major developments in count data travel cost modeling and show how they can be applied to forest-based recreation.

In contrast to earlier HCK modeling that used ordinary least-squares (OLS) regression, count data models emphasize the non-negative, integer nature of data on the number of trips taken and are most useful when the counts (per person) are small. This is often the case with forest recreation, such as backcountry trips or adventure activities, which most people participate in only a few times a year. Although the normal distribution is a good approximation of the Poisson distribution (which is sometimes called the "law of rare events") if the mean of the distribution is large, the normal distribution provides a poor approximation of the Poisson for small mean values. This is due to the skewness of the Poisson distribution (Kalbfleisch 1985). Count data estimators place positive probability only on possible, discrete events. OLS estimators can place positive probability on fractional and negative events (Creel and Loomis 1990). Thus, for small counts of recreational trips, a count data distribution is more likely to represent the true data-generating process than is a normal distribution.

The remainder of the chapter is divided into four sections. The next section provides a brief history of the development of the count data travel cost model. The second section outlines the basic theoretical and empirical issues that must be considered in designing a count data analysis of recreation demand. In section 3, we use count data models to estimate the value of rain forest protection in Brazil. In the final section, we summarize the chapter and discuss the needs for future research in count data recreation demand modeling.

1. A BRIEF HISTORY OF COUNT MODELS IN TRAVEL COST ANALYSIS

Over the past two decades, applied econometricians have paid increasing attention to estimation and testing of count data models. Early econometric analyses of count data models include the effects of research and development on patents issued (Hausman et al. 1984) and the relationship between urban air quality and respiratory illness (Portney and Mullahy 1986).

Not long after these foundational studies, it was realized that data collected for the HCK class of travel cost models were amenable to analysis using count data models. To the authors' knowledge, Shaw (1988) was the first to apply count data models to recreation demand. Shaw recognized that recreation data collected on site are truncated and may suffer from endogenous stratification (people who frequently visit a site are more likely to be sampled than people who rarely visit), and that failure to correct for these problems leads to biased estimates of population parameters. Shaw's estimator is presented in section 2.4.

Economists are always cognizant of the need for empirical models to be consistent with an underlying theoretical foundation. In 1993, Hellerstein and Mendelsohn provided such a foundation for count data travel cost models. They realized that on any choice occasion, the decision of whether or not to take a trip to a specific site can be modeled using a binomial distribution and, as the number of choice occasions increases throughout a recreational season, the binomial distribution asymptotically converges to a Poisson distribution.

During the 1990s, Poisson and negative binomial count data models (presented in sections 2.2 and 2.4) were estimated for a variety of recreational resources. Hellerstein (199 1) estimated count data models for trips to the Boundary Waters Canoe Area and showed how models could be estimated using aggregate (zonal) data. Creel and Loomis (1990) tested a variety of Poisson and negative binomial estimators and found that count data models were more appropriate for estimating and predicting the demand for deer hunting in California than were OLS and nonlinear least-squares estimators. Yen and Adamowicz (1993) evaluated the statistical properties of welfare measures computed using count data models of the demand for hunting bighorn sheep and suggested caution when evaluating consumer surplus measures derived from truncated estimators (which are used for analysing on-site data). Englin and Shonkwiler (1995) developed a truncated, endogenously stratified negative binomial model and used it to estimate long-run demand for overnight hikes in the Cascade Mountains. Ovaskainen et al. (2001) used the Englin-Shonkwiler estimator to model the demand for forest recreation trips in Finland.

A recent development in count data modeling is based on the realization that observations of zero trips may be generated either by people who are not in the market (they would not take a trip at any positive price) or by people who are in the market but did not take a trip during the observation period (the price faced in the observation period was too high). Zero-inflated models (also referred to as augmented count or double-hurdle models) have not been as frequently applied as other count data models, probably because they require samples of the entire population. Shonkwiler and Shaw (1996) clarified the nature and interpretation of hurdle count data models, including single-hurdle selection-type models and zero-inflated models. Haab and McConnell (1996) presented zero-inflated models for beach trips; Shaw and Jakus (1996) showed how to estimate a zero-inflated model of the demand for rock climbing; and Gurmu and Trivedi (1996) estimated the demand for lake recreation using zero-inflated models. This approach is discussed in section 2.5.

2. THEORY AND EMPIRICAL ANALYSIS

2.1 Linear Exponential Demand and Welfare Estimates

Unlike demand functions based on the normal distribution, expected values in demand functions based on count data models are restricted to be non-negative. A functional form that guarantees positive mean values is the linear exponential (semi log) demand function:

$$E[Q_i] = e^{X_i\beta} \qquad 19.1$$

where $E[Q_i]$ is the expected number of visits to a site by individual i, X_i is a vector of observations on independent variables associated with individual i (including the travel cost), and the βs are parameters to be estimated. The specification in equation 19.1 makes a clear distinction between the functional form of the demand curve and the distributional assumptions used to obtain estimates of the demand function parameters. For count data models, the demand function is linked with a count data distribution by the relationship:

$$\lambda_i = E[Q_i] = e^{X_i\beta} \qquad 19.2$$

where λ_i is the mean of the count data distribution (for individual i).

Equation 19.1 represents expected, not actual, demand, because the equation errors associated with individual heterogeneity do not enter the expression. Therefore, Marshallian consumer surplus in a count model is computed for the typical consumer.

As usual, consumer surplus is found by integrating the area under the demand curve from a lower price p_0 to an upper price p_1. For the linear exponential demand function, this integration yields the expression:

$$CS_i = \int_{p_0}^{p_1} \lambda_i dp = \frac{\lambda_i(X_i^{p_1}) - \lambda_i(X_i^{p_0})}{\beta_{tc}} \qquad 19.3$$

where β_{tc} is the parameter estimate on travel cost, X_i^{p1} is the X_i vector substituting p_1, and X_i^{p0} is the X_i vector substituting p_0. If p_1 is the choke price (the price at which the expected number of trips equals zero), then consumer surplus is written as:

$$CS_i = \int_{p_0}^{p_1} \lambda_i dp = \frac{-\lambda_i(X_i^{p_0})}{\beta_{tc}} \qquad 19.4$$

because $\lambda_i(X_i^{p1}) = 0$. Equation 19.3 would be useful if the analyst wanted to estimate the change in surplus associated with a marginal increase in price, such as an increase in user fees. Equation 19.4 provides an estimate of total consumer surplus associated with the site. By dividing total consumer surplus by the number of trips (λ_i), it is easily seen that Marshallian consumer surplus per trip is $-1/\beta_{tc}$.

It is important to recognize that welfare estimates should only be applied to the sample frame from which the X_is are drawn. If the estimator used to obtain the βs recovers the population parameters, then consumer surplus and latent demand (the desired number of trips) are found by simulating the demand equation using population means for the independent variables. This procedure obtains a consumer surplus estimate for the typical member of the population only if the population parameters were accurately recovered from the available sample.

To test hypotheses about consumer surplus estimates (such as the hypothesis that consumer surplus associated with a project exceeds project cost), it is necessary to obtain estimates of variance. Englin and Shonkwiler (1995) showed that the second-order Taylor series approximation of the variance of consumer surplus associated with linear exponential demand is:

$$Var\left(\frac{1}{\beta_{tc}}\right)=\frac{V}{\beta_{tc}^4}+2\frac{V^2}{\beta_{tc}^6} \qquad 19.5$$

where V is the variance of β_{tc}. Of course, if estimates of exact changes in welfare are desired, then Hicksian measures are required. Bockstael et al. (undated) determined the compensating variation (CV) and equivalent variation (EV) formulas to be:

$$CV=\frac{1}{\beta_y}\ln(1+\frac{\lambda\beta_y}{\beta_{tc}}) \qquad 19.6$$

and

$$EV=-\frac{1}{\beta_Y}\ln(1-\frac{\lambda}{\beta_{tc}}) \qquad 19.7$$

where β_y is the coefficient on the income variable, and β_{tc} is the coefficient on the travel cost variable. Englin and Shonkwiler (1995) provide a method for calculating variances around these Hicksian welfare measures.

2.2 Econometric Analysis of Single-Site Count Models

Two versions of single-site count models have been developed. The first version applies when data are available only for a specific site. These models are based on either data for individual visitors or zonal data. The variation in prices needed to estimate the demand curve is obtained by pooling different individuals who face different travel costs. Other explanatory variables are included in the specification to control for variation in socio-economic characteristics (such as income).

Studies based on a specific site are unusual. In the second version of single-site models, data are pooled across individuals and sites. This approach imposes the restriction that parameters are the same across all of the pooled sites. If this restriction is accepted and individuals and sites are both pooled, then the independent variables include the characteristics of the sites as well as the individuals. Examples of this approach include Creel and Loomis (1990, 1992), Englin and Shonkwiler (1995), and Ovaskainen et al. (2001).

Single-site count data models are usually estimated using either the Poisson or the negative binomial distribution. The probability density function (*pdf*) for the Poisson is a one-parameter distribution (the mean equals the variance) and is written as:

$$\Pr(Q_i = q_i) = \frac{e^{-\lambda_i}\lambda_i^{q_i}}{q_i!} \qquad 19.8$$

where q_i is a non-negative integer, and the log-likelihood function for a sample of size n is given by:

$$\ln L = \sum_{i=1}^{n}[-e^{X_i\beta} + q_i X_i\beta - \ln q_i!] \qquad 19.9$$

(recall that $\lambda_i = exp(X_i\beta)$). Parameter estimates are obtained by maximizing the log-likelihood function.

The mean-variance equality restriction of the Poisson model has been viewed as its major limitation. One way to account for over-dispersion (variance > mean) in count data is to include a stochastic variable (ε) in equation 19.2 that accounts for heterogeneity across people and allows $\lambda_i = exp(X_i\beta + \varepsilon_i)$ to vary according to a specific probability law. Cameron and Trivedi (1986) show that if $exp(\varepsilon)$ follows a gamma (Γ) distribution, then the compound count data generation process follows a negative binomial distribution. The *pdf* for the negative binomial distribution is:

$$\Pr(Q_i = q_i) = \frac{\Gamma\left(q_i + \frac{1}{\alpha}\right)}{\Gamma(q_i+1)\Gamma\left(\frac{1}{\alpha}\right)}(\alpha\lambda)^{q_i}(1+\alpha\lambda)^{-(q_i+\frac{1}{\alpha})} \qquad 19.10$$

where I/a is a dispersion parameter. Other forms of the dispersion parameter are possible (Cameron and Trivedi 1986). The distribution in equation 19.10 has conditional mean λ_i and conditional variance $\lambda_i (1 + \alpha\lambda_i)$. Since $\lambda_i > 0$ and a > 0, it is clear that the variance is greater than the mean. If the dispersion parameter is not different from zero, the negative binomial model reduces to the Poisson. Software is available for estimating the parameters of both the Poisson and the negative binomial models.

2.3 Practical Issues in Model Specification

Until now this chapter has focused on what may be called the science of recreation count modeling. These tools, while powerful, must be implemented in a manner appropriate to a given context. A number of practical issues arise related to model specification.

To see the question of model specification and welfare estimation, return to the formula for consumer surplus in equations 19.3 and 19.4. Two methods for incorporating trip attributes in the welfare estimates are possible. First, changes in X affect total welfare, and a shift $(X_1 - X_2)$ provides a measure of the associated change in consumer surplus. Englin and Shonkwiler (1995) used this approach to derive the long-run shift in hiking values from demographic shifts over a four-decade period. This method could also be used to derive impacts resulting from changes in site characteristics if that information were included in the Xs. Notice, however, that the change in consumer surplus is entirely driven by changes in visitation.

In some situations, this simple approach may be unsatisfactory. Suppose one were interested in estimating the impact of clearcuts on the economic welfare of hikers. Both the number of trips and the consumer surplus per trip would be affected. A practical remedy is the varying parameter model where the slope of the demand curve is a function of the level of characteristics (Vaughan and Russell 1982). This is accomplished by adding a term that includes the travel cost interacted with the characteristic level. Total consumer surplus in this model becomes:

$$CS_i = \frac{-\lambda_i(X_i^{p_0})}{\beta_{tc}(X_i)} \qquad 19.11$$

where $\beta_{tc}(X_i)$ indicates that the parameter estimate on travel cost is a function of site quality. In general, one simply interacts travel cost and the characteristics linearly, but other specifications are possible.

A second related issue is the measurement of forest ecosystem attributes in a pooled-site model. Given that GIS data are available in many areas, the researcher has several options. Consider the evaluation of hiking trails that pass through different ecotypes where trails cover major changes in elevation. The simplest alternative is to measure the ecotype as present or absent along the trail (Englin and Shonkwiler 1995). A second approach is to use the total area of forest type that a trail goes through, or the total length of the trail that passes through a given ecotype (Pendleton et al. 1998). A third alternative is to use a latent characteristics model to construct bundles of attributes that represent the holistic quality of ecotypes (Pendleton and Shonkwiler 2001). The approach taken will depend, among other things, on the quality of the data available and the variation in the relevant ecosystems. Focus groups and other scoping methods can indicate which of these measures is most relevant to recreationist decision-making.

2.4 On-Site Survey Count Data Estimators

Recreation demand models are often estimated using survey responses collected from on-site samples of visitors. This is because it is generally less expensive to collect data on site than to collect data from the general population. However, it is useful to be able to estimate population parameters from on-site (truncated) samples so that total demand and value can be computed. This is accomplished by adjusting the untruncated models (Grogger and Carson 1991).

The *pdf* of a variable truncated at zero (i.e., zeros are not observed) is simply the untruncated *pdf* $f(q_i)$ divided by the area under $f(q_i)$ where $q_i > 0$. This guarantees that the area under the truncated *pdf* equals 1. For the Poisson distribution, the probability that q_i exceeds zero is $(1 - exp(-\lambda_i))$, and, dividing the expression in equation 19.8 by this probability, the conditional probability for a zero-truncated model is:

$$\Pr(Q_i = q_i) = \frac{e^{-\lambda_i}(\lambda_i)^{q_i}}{q_i!\left[1 - e^{-\lambda_i}\right]} \qquad 19.12$$

This procedure can also be used to obtain the conditional probability for the negative binomial distribution (for example, see Creel and Loomis 1990).

Shaw (1988) recognized that, in addition to truncation at zero, on-site samples are endogenously stratified. That is, people who visit a site often are more likely to be sampled than are people who visit infrequently. He showed that the on-site *pdf* of the i^{th} person in the population is the product of the untruncated *pdf* and the variable q_i/λ_i, which is the ratio of actual trips to expected number of trips for the representative individual with characteristics X_i:

$$\Pr(Q_i = q_i) = \frac{e^{-\lambda_i}(\lambda_i)^{q_i - 1}}{(q_i - 1)!} \qquad 19.13$$

A comparison of equations 19.13 and 19.8 shows that population parameters for the Poisson model, controlling for truncation and endogenous stratification, can be estimated from an on-site data sample by replacing q_i with $(q_i - 1)$. Unfortunately, this convenient result does not hold for the negative binomial model. As shown by Englin and Shonkwiler (1995), the on-site sample's negative binomial density function (found by multiplying equation 19.10 by the ratio q_i/λ_i) is:

$$\Pr(Q_i = q_i) = \frac{q_i \Gamma(q_i + \frac{1}{\alpha}) \alpha_i^{q_i} \lambda_i^{q_i - 1} [1 + \alpha_i \lambda_i]^{-\left(q_i + \frac{1}{\alpha}\right)}}{\Gamma(q_i + 1)\Gamma\left(\frac{1}{\alpha}\right)}. \qquad 19.14$$

While equation 19.13 can be estimated in standard packages, the likelihood function associated with equation 19.14 must be programmed.

2.5 Population Samples, the Participation Decision, and Zero-Inflated Models

Prior research has shown that truncated count data models do not always provide good estimates of population parameters, and substantial benefits may be gained by collecting information on nonparticipants (Yen and Adamowicz 1993). It is logical that the most direct way to estimate population parameters is to collect trip data from a sample of the population. These data would include information on people who did not take recreation trips to the site(s) of interest. Recent research has shown that modeling the participation decision can increase the efficiency of parameter estimates and provide planners with information about the segmentation of recreation markets (Haab and McConnell 1996).

The zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models are more general than the Poisson and negative binomial models in that they relax the restriction that an identical process generates both the zeros and the positive integers. Recreation data collected from users and nonusers of a resource provide two kinds of information: (1) whether or not to participate, and (2) the quantity demanded conditional on the participation decision. Poisson and negative binomial count data models do not extract information about the participation decision from the zeros in the data but treat the zeros as being generated by the same process that generates positive observations.

Instead of a single data-generation process, the ZIP/ZINB models consider that (1) $q_i \sim 0$ with probability p_i, and (2) $q_i \sim$ Poisson or negative binomial with probability $1 - p_i$. For the Poisson model, this implies that:

$$q_i = 0 \quad \text{with probability} \quad p_i + (1 - p_i)e^{-\lambda_i}$$
$$q_i = k \quad \text{with probability} \quad (1 - p_i)\frac{e^{-\lambda_i}\lambda_i^k}{k!} \qquad 19.15$$

where $k = 1,2,3,...$ are positive integers, and $exp(-\lambda_i)$ is the Poisson probability of taking zero trips. Note that in equation 19.15, zero trips can be generated by both a binomial process (for people not in the market) and a Poisson process (for people in the market who took zero trips). This later expression, $(1 - p_i)exp(-\lambda_i)$, represents the probability of a corner solution by potential users.

The (binary) recreation participation decision can be specified as a logistic model:

$$\log\left(\frac{1-p_i}{p_i}\right) = Z_i\gamma \qquad 19.16$$

where γ is a vector of participation-decision parameters, and Z_i is a vector of explanatory variables that may or may not share variables with X_i. Expected consumer surplus per year is estimated by:

$$CS_i = (1 - p_i)\left(\frac{-\lambda_i(X_i^{p_0})}{\beta_{tc}}\right) \qquad 19.17$$

where β_{tc} is the parameter estimate on the travel cost variable. Expected consumer surplus per trip remains (– Z/&J.

2.6 Demand System Analysis

Demand system analysis derives from the realization that there may be several sites that have related demand functions. If so, partial equilibrium analysis must account for multiple sites. Systems of count data demand functions can be motivated by recognizing that the limiting distribution of a multinomial distribution (where an individual chooses where to recreate from a choice set containing multiple sites) is a system of independent Poisson distributions (von Haefen and Phaneuf 2002).

In a demand system, the prices for sites and their substitutes change simultaneously. In the case of n sites being in the partial equilibrium, the demand system can be written as:

$$\ln(q_{ij}) = \alpha_j + \sum_{j=1}^{n} \beta_j p_{ij} + \gamma_j m_i \qquad 19.18$$

where q_{ij} is the number of trips taken by individual i to site j, p_{ij} are travel costs facing individual i for trips to site j, m_i is individual i's income, and α_j, β_j, and γ_j are parameters to be estimated.

Given a set of demand functions of the form shown in equation 19.18, an important question is whether they can be integrated back to the expenditure function (for example, see Bockstael et al. 1991) or, through inversion, the indirect utility function. LaFrance (1990) showed that if the linear exponential demand functions are treated as an incomplete demand system, the associated partial utility function could be recovered. Assuming no income effects, the partial indirect utility function consistent with equation 19.18 is:

$$m - \left(\frac{\alpha_1}{\beta_{11}}\right) e^{\sum_{i=1}^{k} \beta_{ii} p_i} - \sum_{i=k+1}^{n} \left(\frac{\alpha_i}{\beta_{ii}}\right) e^{\beta_{ii} p_i} \qquad 19.9$$

The conditions that a system of semi logarithmic demand functions must fulfill to form an integrable demand system have been well documented. Empirically, the conditions are simply restrictions on the relationships between the intercept, cross-price effects, and the income effect in the model. The intercept restriction is:

$$\alpha_j \geq 0 \qquad 19.20$$

or non-negativity, where α_j is the intercept for the j^{th} site. A second restriction is that the income effect (β) is restricted to be the same across equations. A final restriction is that the Marshallian cross-price effects are all zero.

For a Poisson count demand system, the likelihood function is simply the product of the single-site demands shown in equation 19.8. The joint likelihood function is:

$$\prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{q_i}}{q_i!} \qquad 19.21$$

where n is the number of sites and the latent quantity demanded for a given site i, λ_i, is $exp(X_i\beta)$. Equation 19.21 is maximized subject to the restrictions described above.

A good example of the application of a Poisson demand system to forest recreation is the study by Englin et al. (1998). They focused on the impact of exchange rates on demand for backcountry canoeing in four wilderness parks in Canada. Because visitors from the United States use these parks, a shift in the exchange rate not only changes the prices of all parks but also changes the relative prices of parks within the system to different consumers. This is because the proportion of the travel costs that occur within the United States versus Canada differs across individuals. For Canadian visitors, the Canadian costs comprise total travel cost. For some visitors, such as visitors from the north central United States, the Canadian costs comprise a large proportion of total travel costs. However, for other visitors, such as from the southern United States, Canadian costs are a tiny proportion of total costs.

Measuring welfare in a setting where relative price shifts are simultaneously distributed across several substitute sites requires a Hicksian framework for analysis. As pointed out by Englin et al. (1998), the restriction that Marshallian cross-price effects must equal zero does not mean that the Hicksian cross-price effects are zero. The Hicksian cross-price effects can be calculated as:

$$s_{ijk} = q_{ij} \frac{\partial q_{ik}}{\partial m_r} = \gamma q_{ik} q_{ij} \qquad 19.22$$

where s_{ijk} is the Hicksian substitution effect between sites j and k for individual i, and the q's are quantities of trips to the sites in the system by individual i. This is simply an application of the Slutsky formula. The Hicksian cross-price effects are symmetric (i.e., $s_{jk} = s_{kj}$) for any individual i. If no individual takes trips to specific pairs of sites, then some cross-price effects are zero. For example, Englin et al. (1998) found that the cross-price effects for the most remote and most developed parks were zero, suggesting that subsets of parks provide opportunities for different types of recreationists.

3. APPLICATION: RAIN FOREST VALUATION

To demonstrate how count data methods can be applied to real world forestry problems, we present a case study of tropical rain forest protection in southern Brazil. There are relatively few studies of the *in situ* value of tropical rain forests. This example provides estimates of the recreational value of an area designated by UNESCO (United Nations Educational, Scientific and Cultural Organization) as the Lagamar Biosphere Reserve and, to our knowledge, represents the first application of a count data model to recreation in Brazilian protected areas.

3.1 Background

The Atlantic Coastal Forest stretches for more than 3,000 kilometres along the coast of Brazil. This ecosystem is rich in biological diversity and endemic species and is considered to be one of the most endangered ecosystems in the world. The largest remaining contiguous area of this forest type occurs in southern Brazil in the Environmental Protection Area (Area de Proteção Ambiental, or APA) of Guaraqueçaba.

The forests within and surrounding the APA have been protected in large part due to the area's isolation, which has limited tourism and other forms of economic development. Beyond the mountains and the bay that form the boundary of the region lie some of Brazil's largest and most economically developed cities, which are potential sources both of deforestation pressure and of tourists.

At the time of the study, recreation of any type was very limited in the APA due to difficult accessibility. The APA could be reached by following a dirt road, much of it in poor repair, for more than 60 kilometres. Adventure tourism in the immediately surrounding area included mountaineering, hiking, and camping in the protected Marumbi Area (Serra do Mar mountains) and primitive beach recreation on the protected Ilha do Mel Ecological Station. Immediately outside of these primitive areas, mass tourism was occurring at heavily developed beaches along the coast.

3.2 Sampling Methods and Data

Data for this case study were collected from Brazilian tourists at adventure tourism sites and other locations. On-site interviews were conducted at six adventure tourism sites in the study area. The off-site sample was drawn from tourists at 25 popular locations outside of the adventure tourism areas. All respondents were asked to indicate how many trips they had taken to each site during the past year. Respondents were also asked where they lived, what recreational activities they participated in, what investments the government should make (if any) in public recreation, and some socio-economic questions. Complete records were obtained for 143 people on-site and 337 people off-site. Following Creel and Loomis (1990), trips to adventure tourism sites (the APA, Marumbi, and Ilha do Mel) were pooled and treated as a single site. This approach imposes the assumption that parameters are the same across the pooled sites.

Distances were computed from origin-destination data using the official Brazilian road atlas. Travel cost (COST) was estimated by multiplying $0.15/mile times the round-trip distance. Income was defined as monthly household income and was included in the model in logarithmic form (LINC). Numerical scales were created identifying the number of adventure activities (such as mountaineering and hiking) and passive activities (such as sightseeing and picnicking) people participated in (ACTIVE, PASSIVE). Socio-economic variables in the model are respondent age (AGE) and gender (SEX). Finally, a dummy variable indicates the importance of paving the access road into the region (ACCESS). While paving the road would reduce the time and effort required to access the recreation sites, it would also change the character of the region by promoting economic development.

3.3 Results

We estimated a variety of models including Shaw's Poisson and the Englin-Shonkwiler negative binomial (using on-site data) and the ZIP model (using off-site data). Because the dispersion parameter was not significant in the zero-inflated negative binomial model using off-site data, estimates for this model are not reported here.

As shown in table 19.1, each model had the expected negative parameter estimate on the travel cost variable, which is consistent with a downward-sloping demand curve. Also consistent across the models, men and those who participated in more adventure activities took more trips. The dispersion parameter was marginally significant in the Englin-Shonkwiler negative binomial model, suggesting that it was more suitable than Shaw's Poisson for these data.

The ZIP model reveals information about the decision of whether or not to participate in adventure tourism. As can be seen in the bottom panel of table 19.1, those who participated in more adventure-related activities, those who did not think that access to the adventure sites should be improved, and those with lower incomes were more likely to visit the adventure tourism sites. Using equations 19.15 and 19.16 and the parameter estimates from the ZIP model, we estimated that 55% of the population would not take a trip to the adventure sites. Of the estimated 45% of the population who were in the market, only one-third (or 15% of the population) were likely to have taken a trip in the survey period. An estimated two-thirds of the people in the market did not take a trip but were located at the demand comer point. This result suggests that there is a large potential demand for primitive recreation sites in this region of Brazil.

Table 19.1. Parameter estimates for on-site and off-site count data models in southern Brazil

Variable	Shaw's Poisson (St. Err.)	Englin-Shonkwiler Negative Binomial (St. Err.)	Zero-inflated Poisson (St. Err.)
Trips equation	- On-site -		- Off-site -
Constant	1.512**	1.50	-0.761
	(0.766)	(1.064)	(0.737)
COST	-0.016***	-0.022*	-0.007***
	(0.004)	(0.013)	(0.002)
ACTIVE	0.360***	0.402***	0.190**
	(0.049)	(0.063)	(0.080)
PASSIVE	0.254***	0.308***	-0.026
	(0.096)	(0.119)	(0.134)
AGE	-0.022**	-0.025*	-0.005
	(0.011)	(0.015)	(0.006)
SEX	-0.426**	-0.570**	-0.611**
	(0.176)	(0.228)	(0.287)
LINC	-0.089	-0.063	0.193*
	(0.098)	(0.129)	(0.106)
σ (overdispersion)	—	0.151*	
		(0.084)	
Participation equation			
Constant	—	—	3.121
			(1.920)
ACTIVE	—	—	0.516**
			(0.216)
LINC	—	—	-0.547**
			(0.284)
ACCESS			-1.26**
			(0.542)
N	143	143	337
Predicted trips/year	1.10	1.45	0.39
CS/trip [a]	$62.50	$45.45	$142.86
CS/year [a]	$58.13	$65.91	$55.71

[a] CS = Consumer Surplus

*** = significant at 1% level, ** = significant at 5% level, * = significant at 10% level

In comparison, results from the on-site sample indicated that from 67% (using parameters from Shaw's Poisson model) to 77% (using parameters from the Englin-Shonkwiler negative binomial model) of the population were in the market. The discrepancy between market shares estimated using on-site and off-site models is due to the ability of the off-site model to locate the comer point of the recreation demand curve.

Tobias and Mendelsohn (1991), using the same travel cost per kilometer as in our study, estimated consumer surplus per trip to be $35 for domestic tourists visiting a rain forest reserve in Costa Rica. Our estimates, using on-site and off-site data, are higher than reported for Costa Rica. This is not surprising given the proximity of relatively affluent urban areas (such as São Paulo and Curitiba) to our study sites. Taken together, these studies indicate that protection of tropical rain forests can provide significant recreational benefits to citizens in developing countries.

4. CONCLUSIONS

Estimates of the value of outdoor recreation provide policy makers with information that is essential to planning multiple-use management of forests. Count data models are a relatively new addition to recreation demand modeling and focus attention on the nature of the underlying processes that generate data on recreational trips. If the number of recreational trips taken by visitors is small, the true data-generation process cannot be normally distributed. In these cases, OLS models are inappropriate for estimating travel cost demand curves. Count data models that are consistent with the non-negative, integer nature of trip data are required. While a Poisson model is often the starting point for estimation, the potential for over-dispersion should always be evaluated.

Researchers often collect data by sampling recreationists on-site in order to save money on survey costs. If data are collected on-site, adjustments for truncation and endogenous stratification must be made in order to estimate population parameters. However, these estimators are founded on the (generally untested) assumption that identical processes generate positive trips and zero trips in the population.

Although off-site data may be more expensive to collect than on-site data, the gain in information may exceed the incremental cost. That gain includes the ability to jointly model the participation decision along with the number of trips taken. Such zero-inflated models permit the researcher to isolate three types of people in the population: (1) those who would never take a trip, (2) those who would take a trip if the price were low enough, and (3) those who are trip-takers. In cases where there exists a large potential recreation demand, such as our case study in Brazil, this representation should result in more accurate estimates of the recreational value of forests.

5. LITERATURE CITED

BOCKSTAEL, N., W.M. HANEMANN, AND I. STRAND (eds.). Benefit analysis using indirect or imputed market methods, Vol. II. Department of Agricultural and Resource Economics, University of Maryland, College Park (undated).

BOCKSTAEL, N.E., K.E. MCCONNELL, AND I. STRAND. 1991. Recreation. P. 227-270 *in* Measuring the Demand for Environmental Quality, Braden, J.B., and C.D. Kolstad (eds.). Elsevier, New York. 370 p.

BROWN, G., JR., AND R. MENDELSOHN. 1984. The hedonic travel cost method. Rev. Econ. Stat. 66:427-433.

CAMERON, A.C., AND P. TRIVEDI. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. J. Applied Econ. 1:29-53.

CLAWSON, M. 1959. Methods of measuring the demand for and value of outdoor recreation. Resources for the Future Reprint No. 10, Washington DC.

CLAWSON, M., AND J.L. KNETSCH. 1966. Economics of outdoor recreation. Johns Hopkins University Press (for Resources for the Future), Baltimore, MD. 328 p.

CREEL, M., AND J. LOOMIS. 1990. Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. Am. J. Agric. Econ. 72:434-441

CREEL, M., AND J. LOOMIS. 1992. Modeling hunting demand in the presence of a bag limit, with tests of alternative specifications. J. Environ. Econ. Manage. 22: 99-l 13.

ENGLIN, J., AND R. MENDELSOHN. 1991. A hedonic travel cost analysis for valuation of multiple components of site quality: The recreation value of forest management. J. Environ. Econ. Manage. 2 1: 275-290.

ENGLIN, J., AND J.S. SHONKWILER. 1995. Estimating social welfare using count data models: An application to long-run recreation demand under conditions of endogenous stratification and truncation. Rev. Econ. Stat. 77: 104-I 12.

ENGLIN, J., PC. BOXALL, K. CHAKRABOTY, AND D.O. WATSON. 1996. Valuing the impacts of forest fires on backcountry forest recreation. For. Sci. 42: 450-455.

ENGLIN, J., P. BOXALL, AND D. WATSON. 1998. Modeling recreation demand in a Poisson system of equations: An analysis of the impact of international exchange rates. Am. J. Agric. Econ. 80: 255-263.

FREEMAN, A.M. III. 1993. The measurement of environmental and resource values: Theory and methods. Resources for the Future, Washington DC. 516 p.

GROGGER, J., AND R. CARSON. 1991. Models for truncated counts. J. Applied Econometrics. 6:225-238.

GURMU, S., AND P.K. TRIVEDI. 1996. Excess zeros in count models for recreational trips. J. Bus. Econ. Stat. 14: 469-477.

HAAB, T.C., AND K.E. MCCONNELL. 1996. Count data models and the problem of zeros in recreation demand analysis. Am. J. Agric. Econ. 78: 89-102.

HAUSMAN, J.A., B.H. HALL, AND Z. GRILICHES. 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52: 909-938.

HELLERSTEIN, D. 1991. Using count data models in travel cost analysis with aggregate data. Am. J. Agric. Econ. 73: 860-867.

HELLERSTEIN, D., AND R. MENDELSOHN. 1993. A theoretical foundation for count data models, with an application to a travel cost model. Am. J. Agric. Econ. 75: 604-611.

HOLMES, T., B. SOHNGEN, L. PENDLETON, AND R. MENDELSOHN. 1997. Economic value of ecosystem attributes in the Southern Appalachian Highlands. P. 187-190 *in* Proceedings of the Conference on Integrating Social Sciences and Ecosystem Management, H.K. Cordell (ed.). General Technical Report SRS-17, USDA Forest Service, Asheville, NC. 230 p.

KALBFLEISCH, J.G. 1985. Probability and statistical inference. Springer-Verlag, New York. 703 p.

LAFRANCE, J. 1990. Incomplete demand systems and semi-logarithmic demand models. Australian J. of Agric. Econ. 34: 118-131.

OVASKAINEN, V., J. MIKKOLA, AND E. POUTA. 2001. Estimating recreation demand with on-site data: An application of truncated and endogenously stratified count data models. J. Forest Econ. 7: 125-144.

PENDLETON, L.H., B. SOHNGEN, R. MENDELSOHN, AND T. P. HOLMES. 1998. Measuring environmental quality in the Southern Appalachian Mountains. For. Sci. 44: 603-609.

PENDLETON, L.H., AND J.S. SHONKWILER. 2001. Valuing bundled attributes: a latent characteristics approach. Land Econ. 77: 118-129.

PORTNEY, P., AND J. MULLAHY. 1986. Urban air quality and acute respiratory illness. J. Urban Econ. 20: 21-38.

SHAW, DAIGEE. 1988. On-site samples' regression: Problems of non-negative integers, truncation and endogenous stratification. J. Econometrics 37: 211-223.

SHAW, W.D., AND P. JAKUS. 1996. Travel cost models of the demand for rock climbing. Ag. Res. Econ. Rev. 25: 133-142.

SHONKWILER, J.S., AND W.D. SHAW. 1996. Hurdle count-data models in recreation demand analysis. J. Agric. Res. Econ. 21: 210-219.

TOBIAS, D., AND R. MENDELSOHN. 1991. Valuing ecotourism in a tropical rain-forest reserve. Ambio 20(2): 91-93.

VAUGHAN, W.J., AND C.S. RUSSELL. 1982. Valuing a fishing day: An application of a systematic varying parameter model. Land Econ. 58: 451-463.

VON HAEFEN, R.H., AND D.J. PHANEUF. 2002. A note on estimating nested CES preferences for outdoor recreation. Unpublished Working Paper. Department of Agricultural and Resource Economics, North Carolina State University, Raleigh. 18 p.

YEN, S.T., AND W.L. ADAMOWICZ. 1993. Statistical properties of welfare measures from count-data models of recreation demand. Rev. Agric. Econ. 15: 203-215.