Journal of Leisure Research
2008, Vol. 40, No.3, pp. 457–478

Welfare Measurement Convergence Through Bias Adjustments in General Population and On-Site Surveys: An Application to Water-based Recreation at Lake Sevan, Armenia

Craig Meisner and Hua Wang
Development Research Group, The World Bank

Benoit Laplante
Independent Consultant, Montreal, Canada

Abstract

This paper compares household survey with on-site survey data for estimating the access value of a unique natural resource using a single-site travel cost model. The household survey model is adjusted for inflated zero observations for respondents who would not visit the site at any observable positive price. The on-site survey model is corrected for truncation and endogenous stratification, the latter being an adjustment for avidity bias. In an application to recreation at Lake Sevan (Armenia), consumer surplus estimates were not statistically different between the household model and the on-site model when zero-inflation and truncation and endogenous stratification are corrected in the respective models. This leads us to believe that either method can be used to derive a consistent welfare measure of access to a recreational site after appropriate adjustments and corrections are made. These results are somewhat reassuring as the choice between household and on-site surveys is often dictated by time and resource availability.

KEYWORDS: *On and off-site sampling, zero-inflation, count data, endogenous stratification, Armenia.*

Introduction

Travel cost-based demand estimation models rely on actual site visitation data. These data can be collected either on-site or through general population surveys. Each survey type has its own advantages and disadvantages. For example, while general population surveys have the potential to be more broadly representative of a population, they may suffer from known biases such as respondent recall bias if the site in question is infrequently visited or disproportionately zero visits if a large proportion

---

Address correspondence to: Craig Meisner, MSN #MC3-300, World Bank, 1818 H Street, NW, Washington, DC 20433, cmeisner@worldbank.org.

Authors Note: This paper is an outcome of the research project "Measuring the Economic Value of Environmental Protection Projects: Methodology and Application to Armenia's Lake Sevan", Research Grant RF-P070055-RESE, World Bank. The views expressed here are the authors' and do not reflect those of the World Bank, its Executive Directors, or the countries they represent.

of the survey sample is not visiting the site at all. On the other hand, on-site surveys have the advantage of precision regarding the time, date and certainty of visit. However, they run the risk of over-sampling only those in the population who are avid users of the site while not sampling potential participants should constraints to the participation decision be relaxed. We are aware of two studies that directly evaluated the convergence of benefit estimates for the same resource based on data from general population and on-site surveys (Loomis, 2003; Shaw *et al.*, 2003). We extend this literature by evaluating the convergence of demand models based on data from general population and on-site surveys for a resource when adjustments are made for statistical biases unique to each survey mode.

A robust comparison of estimates obtained from each sample requires addressing a number of important statistical issues. In particular, recreation demand measured from a population-based household survey is typically censored due to the observation of a large number of zeros (or non-users of the site). Simply treating all zeros in the sample as users of the site may introduce an upward bias of the demand and welfare measures (Shonkwiler and Shaw, 1996; Haab and McConnell, 1996; Gurmu and Trivedi, 1996).[1] On the other hand, on-site surveys have at least two separate issues. The first is that observations of visitation are truncated at one since it surveys only users at the site, while demand estimation requires observations at zero to establish a choke price. A second issue is related to the users surveyed at the site, namely endogenous stratification. On-site survey data may lead to biased standard errors and welfare measures if the sample is endogenously stratified; i.e., avid users have higher probabilities of being sampled leading to higher trip frequencies being correlated with their characteristics (Shaw, 1988; Englin and Shonkwiler, 1995).

In the case of household surveys, it is possible to resolve the issue by separating the recreation 'participation' decision from the trip 'quantity' decision using sample selection models, thus reducing the bias introduced by non-users of the site (Haab and McConnell, 1996; Gurmu and Trivedi, 1996). In the case of on-site surveys, it is possible to correct for the potential bias by providing adjustments to the distribution function (Shaw, 1988; Englin and Shonkwiler, 1995). Loomis (2003) and Shaw *et al.* (2003) show that after adjusting for truncation and endogenous stratification in the *on-site survey*, the welfare estimates are comparable with the results from the household surveys. However, neither study accounts for the possibility of zero-inflation (or excess zeros) in the household survey sample.

In this paper, we test the hypothesis of whether the household and on-site demand estimation yield similar welfare measures after accounting for the biases discussed above. For this purpose, we construct single-site travel cost models for a household and on-site survey conducted at Lake Sevan, Armenia. The single-site travel cost model is preferred in this case as it facilitates demand estimation for visitation and associated welfare comparisons. The context of the application is also of policy relevance. Lake Sevan is a unique recreational and historically significant resource with no practical substitutes. In the past 50 years, the level of the lake has fallen 18 meters, with severe physical and ecological consequences. In reaction to this, the Government of Armenia has been pursuing a Lake Sevan Restoration Plan that would attempt to restore the

[1] The direction of this bias assumes that a significant proportion of the population may not be users of the site.

lake to its previous potential. However, a comprehensive knowledge of the potential benefits of the Plan is currently lacking, in particular insofar as recreational values are concerned. This paper also contributes to this knowledge.

The household survey consisted of 3,358 households across Armenia, and the on-site survey of 389 tourists recreating at Lake Sevan.[2] Travel cost models were constructed and estimated using travel expenditure and socio-demographic information provided from both surveys. As visitation rates in the household survey contained a large proportion of zeros and the presence of overdispersion in trip frequency, a zero-inflated negative binomial model was estimated.[3] For the on-site survey, two truncated negative binomial models were estimated with and without an adjustment for endogenous stratification (ES).[4]

The remainder of this paper is structured as follows. The next section provides a description of travel cost and count data models utilized in this study along with recommendations of how to remedy several dependent variable issues typically encountered with household and on-site recreational surveys. In Section III, the two surveys are described in more detail. In Section IV, the results of estimation are presented, along with a comparison in expected trip demand and estimated welfare measures. Section V provides a brief summary and discussion of the findings.

Travel Cost Method

Single-site travel cost methods (TCMs) include zonal and individual, while multi-site TCMs include zonal, individual and random utility models (RUM) (Parsons, 2003). Each type of TCM has its own data requirements, objectives, and behavioral and statistical assumptions (Fletcher *et al.*, 1990). 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. Many RUM models emphasize the impact of site quality on recreation demand and are estimated using either multinomial or nested logit models. The outcome is a set of probabilities of the likelihood of visiting a particular site. In our application, we employ a single-site individual TCM to estimate the visitation demand function and associated welfare effects from alternative sampling strategies. In particular, we are not concerned with site attributes, quality or the probability of a choice occasion. Furthermore, as Lake Sevan is unique and offers a wide variety of recreational opportunities for which there are few good substitutes, the single-site TCM specification is appropriate.

In the single-site travel cost model, the decision to recreate is typically modeled as a latent demand, y_i^*, representing the number of trips taken in one year as a function of travel cost (P), features of the current visit (Z) and individual demographic characteristics (X) (Fletcher *et al.*, 1990):

$$y_i^* = f(P_i, X_i, Z_i) \qquad i = 1, 2, \ldots, N \tag{1}$$

[2] The household survey was nationally representative, but over-weighted in the capital city Yerevan, and in the area around Lake Sevan. On-site tourists were from the same population as the household survey, and there was no double-counting of any individual respondent. Also, people from outside Armenia were not included in the sample.
[3] Similar examples using inflation models in TCM can be found in Curtis (2003), Gurmu and Trivedi (1996), Haab and McConnell (1996) and Scrogin et al. (2004).
[4] For similar TCM studies adjusting for truncation and endogenous stratification in TCM see Englin and Shonkwiler (1995), Loomis (2003), Martinez-Espineira and Amoako-Tuffour (2005) and Shaw et al., (2003).

Data for estimating TCMs can be generated either through surveys of the general population or through on-site surveys of actual visitors. Each survey mode results in biases that are unique to the data generation process. General population surveys suffer from observations on non-participation in on-site activities, while on-site surveys suffer from truncation and endogenous stratification. In both sampling methods, the data are non-negative integers requiring count data estimators. Each statistical bias and methodological adjustments will be discussed below.

(i) Count data estimators

An important modeling issue for TCM is non-negative integers observed in individual recreational data (Hellerstein, 1991). Count data models have been shown to provide a better modeling approach than traditional OLS regression procedures (Shaw, 1988; Grogger and Carson, 1991) and are particularly amenable to aggregated socio-economic data (Hellerstein, 1991). Two count distributions that have been widely used are the Poisson and negative binomial distributions. The Poisson distribution function is modeled as follows:

$$\Pr(y_i \mid x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad y_i \geq 0 \tag{2}$$

where $\lambda_i = \exp(x_i\beta)$ is the conditional mean and variance of the number of trips taken y_i, x_i is a vector of covariates and β is a vector of coefficients to be estimated. An undesirable feature of Poisson count models, however, is the assumption that the conditional mean and variance are equal (Yen and Adamowicz, 1993, pg. 205). This is especially problematic in empirical research because conditional variances are typically greater than conditional means in socio-economic data (also known as overdispersion, a form of heteroskedasticity). The presence of overdispersion still allows for consistently estimated means of parameter estimates (Gourieroux *et al.* 1984), but causes the standard errors of these estimates to be biased downward, resulting in erroneous tests of their statistical significance (Cameron and Trivedi, 1986).

The equality of the mean and the variance property of Poisson count models has led to the development of negative binomial models (Hausman *et al.*, 1984). This model allows for overdispersion by combining the Poisson distribution with a gamma distribution and hence allowing for heterogeneity to be gamma distributed.

Let v_i represent an individual's unobserved choice to recreate, with $\exp(v_i)$ following a gamma distribution with mean 1 and variance α. The mean of the resulting distribution is $\lambda_i = \exp(x_i\beta + v_i)$ with the regression model expressed as:

$$\Pr(y_i \mid x_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)}\left(\frac{1}{1+\alpha\lambda_i}\right)^{1/\alpha}\left(\frac{\alpha\lambda_i}{1+\alpha\lambda_i}\right)^{y_i} \quad y_i \geq 0 \tag{3}$$

where $E(y_i) = \lambda_i$ and $Var(y_i) = \lambda_i(1 + \alpha\lambda_i)$, $\Gamma(\bullet)$ is the gamma function and α is the overdispersion parameter. The presence of the α parameter in the calculation of the conditional variance of y, if greater than 0, guarantees that the variance is greater than the mean. As α approaches 0, however, the negative binomial model collapses to the Poisson. Thus, testing for α=0 provides a case for selecting the negative binomial over the Poisson, and indirectly for the presence of overdispersion.

(ii) Zero-inflated adjustments

Data collected off-site through general population surveys may lead to observing a large number of zero participation in on-site activities. In the utility maximization framework, observing a zero implies that the individual is currently at some choke price consuming zero trips. If the current "market" price were to fall below the choke price, the individual would be expected to consume a positive number of trips. However, one may also observe a zero if for some reason (such as age, health-related reasons, etc.) services from the site would never enter an individual's utility function (Haab and McConnell, 1996). Thus, there is an important distinction between observing zeros for those who are potential participants and for those who are non-participants. Standard count data models such as the Poisson or negative binomial assume that all individuals surveyed are potential users of the site, and that the same variables influence all potential users similarly (Haab and McConnell, 1996). Put another way, this says that all observations are coming from the same data generating process, when in fact they are not—one is about the participation decision and another is about the amount. The standard Poisson process is not capable of distinguishing these two decisions.

To account for the participation issue, we consider two augmented count data models which account for the presence of a large number of zeros - the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) (Mullahy, 1986; Lambert, 1992; Greene, 1994; Haab and McConnell, 1996).[5] By distinguishing between participants and non-participants, the zero observations may contain valuable information, and a gain in efficiency will be achieved by including all of the observations (Haab and McConnell, 1996, pg. 90). Empirically, zero-inflated count models change the mean structure to allow zeros to be generated by two distinct processes, one for the participation decision (logit or probit) and one for the mean number of trips given participation is positive (count model).[6] By expanding the standard count model to allow for individual-specific characteristics, which may keep an individual from entering the recreation market, one can separate factors which influence the participation decision from those that influence the decision about the number of trips to take to a recreation site (Haab and McConnell, 1996). In estimation, the ZIP model allows for overdispersion in the Poisson data generating process by allowing a mass of zero observations independent of the true Poisson process.

The distribution function for the ZIP model is:

$$\Pr(y_i \mid x_i) = \begin{cases} P_i + (1-P_i)e^{-\lambda_i} & \text{if } y_i = 0, \\ (1-P_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & \text{otherwise.} \end{cases} \tag{4}$$

where $E(y_i) = (1 - P_i)\lambda_i$, $Var(y_i) = (1 - P_i)(1 + P_i\lambda_i)\lambda_i$, and P_i is the probability of zero visitation, with mean $\lambda_i = \exp(x_i\beta)$. Note that in this formulation, zeros can occur in

[5] Zero-inflated models are quite similar to hurdle models since either can separate the data generating process into two decisions, as well as accommodate excess zeros. For a more detailed comparison of the models see Gurmu and Trivedi (1996).

[6] The zero-inflated models also differ from the Heckman continuous two-stage model as they allow for zero observations in the second stage of the decision process (in the negative binomial model for trip frequency).

either the binomial process (when $y_i = 0$) or the Poisson process (when $y_i \geq 1$), since $\exp(-\lambda_i)\lambda_i^0/0! = \exp(-\lambda_i)$. Again, λ_i can be modeled as $\exp(x_i\beta)$, and P_i as $g(z_i\gamma)$, where γ is a vector of participation-decision parameters and z_i is a vector of explanatory variables that may or may not be the same as those for the quantity decision, x_i. The function g(•) can be modeled using either a logit or probit function as they both give similar results. In the presence of dependent variable overdispersion (variance>mean), the participation decision can be similarly decomposed in a zero-inflated negative binomial model as:

$$\Pr(y_i \mid x_i) = \begin{cases} P_i + (1-P_i)\left(\dfrac{1}{1+\alpha\lambda_i}\right)^{1/\alpha} & \text{if } y_i = 0, \\[2ex] (1-P_i)\dfrac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)}\left(\dfrac{1}{1+\alpha\lambda_i}\right)^{1/\alpha}\left(\dfrac{\alpha\lambda_i}{1+\alpha\lambda_i}\right)^{y_i} & \text{otherwise.} \end{cases} \qquad (5)$$

where $E(y_i) = (1 - P_i)\lambda_i$ and $\mathrm{Var}(y_i) = (1 - P_i)[1 + \lambda_i(\alpha + P_i)]\lambda_i$. As before, the presence of the α parameter in the calculation of the conditional variance of y (if greater than 0), guarantees that the variance is greater than the mean, and testing for α=0 provides a case for selecting the negative binomial over the Poisson.

The flexibility of modeling the participation decision in this manner has led to a number of interesting applications in recreational demand analysis, including beach trips (Shonkwiler and Shaw, 1996; Haab and McConnell, 1996), rock climbing (Shaw and Jakus, 1996), lake recreation, (Gurmu and Trivedi, 1996), water-based recreation (Curtis, 2003), and angling site choice (Scrogin *et al.*, 2004).

(iii) Truncation and endogenous stratification

Interview surveys conducted on-site obviously avoid the non-participation issue, but as the dependent variable y_i is strictly non-zero, the distribution of trips is truncated. By not accounting for the truncation, estimates will be biased and inconsistent since the conditional mean is misspecified (Shaw, 1988; Creel and Loomis, 1990; Grogger and Carson, 1991; Yen and Adamowicz, 1993; Englin and Shonkwiler, 1995). In addition, because the sample is on-site, there is a higher likelihood of intercepting a person whose characteristics are correlated with higher trip frequencies, or what is known as 'endogenous stratification' in sampling. The implication is that the truncated demand relationship measures only those with smaller error terms.[7] For the measurement of welfare this is important since consumers surplus estimates will be biased upwards as they represent the preferences of avid recreationists disproportionately.

The simultaneous effect of truncation and endogenous stratification was first explored by Shaw (1988) in the case of the Poisson distribution and extended by Englin and Shonkwiler (1995) to the negative binomial distribution. The procedure involves weighting individual observations by the inverse of the expected value of trips. As-

[7] The error terms are smaller from measuring only avid users of the site.

suming that the density function of the i^{th} person in the population is $f(y_i^* | x_i)$, Shaw (1988) shows that the density function of the same person in the on-site population is:

$$\Pr(y_i \mid x_i) = \frac{y_i f(y_i \mid x_i)}{\sum_{t=1}^{\infty} t \cdot f(t \mid x_i)} \tag{6}$$

If the conditional density $f(y_i^* | x_i)$ is chosen to be Poisson with the location parameter λ_i, then the on-site sample's density function is:

$$\Pr(y_i \mid x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i - 1}}{(y_i - 1)!} \tag{7}$$

where $E(y_i | x_i) = \lambda_i + 1$ and $Var(y_i | x_i) = \lambda_i$. Defining $w_i = y_i - 1$, the standard Poisson model can be estimated, substituting w_i for y_i in (7) above.

In the presence of overdispersion in count data models, the equality of the mean and variance is violated and thus the negative binomial model is preferred with the following density function (Englin and Shonkwiler, 1995):

$$\Pr(y_i \mid x_i) = \frac{y_i \Gamma(y_i + 1/\alpha_i)}{\Gamma(y_i + 1)\Gamma(1/\alpha_i)} \left(\frac{1}{1 + \alpha_i \lambda_i} \right)^{1/\alpha_i} \left(\frac{\alpha_i}{1 + \alpha_i \lambda_i} \right)^{y_i} \lambda_i^{y_i - 1} \tag{8}$$

where $E(y_i | x_i) = \lambda_i + 1 + \alpha_i\lambda_i$ and $Var(y_i | x_i) = \lambda_i(1 + \alpha_i + \alpha_i\lambda_i + \alpha_i^2\lambda_i)$.[8] Following Englin *et al.* (2003) and Ovaskainen *et al.* (2001), we also make the simplifying restriction that the overdispersion parameter be the same across all individuals ($\alpha_i = \alpha$). As the specification in (8) cannot be transformed into any simpler form as in the case of the truncated Poisson, the likelihood function must be programmed directly into a likelihood maximization routine.[9] The log likelihood function used in this context is:

$$\ln L = \sum_{i=1}^{N} \begin{bmatrix} \ln y_i + \ln(\Gamma(y_i + 1/\alpha)) - \ln(\Gamma(y_i + 1)) - \ln(\Gamma(1/\alpha)) + \\ y_i \ln \alpha + (y_i - 1)\ln \lambda_i - (y_i + 1/\alpha)\ln(1 + \alpha\lambda_i) \end{bmatrix} \tag{9}$$

Defining λ_i as the expected number of person-day-trips individual i takes to the site in a year, the empirical demand relationship can be defined as:

$$\lambda_i = \exp(X_i \beta + \varepsilon_i) = \exp(\beta_p p_i + x_i\gamma + \varepsilon_i) \quad i = 1,\ldots,n \tag{10}$$

where β is a K x 1 vector of parameters, X_i is a 1 x K vector of explanatory variables for individual i, p_i is the travel cost for individual i to the site, x_i is the 1 x K −1

[8] See Cameron and Trivedi (1990) or Cameron and Trivedi (2001, p. 336) for details on the LR test for overdispersion.
[9] The likelihood function was programmed into a maximum likelihood routine in STATA.

vector of explanatory variables after p_i is subtracted from X_i, β_p is the parameter on travel cost, and γ is the remaining vector of parameters corresponding to x_i.

(iv) Welfare measures

The benefit (consumer surplus) of access to the site is defined as the area under the estimated Marshallian demand curve specified in (10) and above the current price level. By integrating the demand function from average travel cost (price, P_o) to the choke price (P_c), we calculate expected consumers surplus as:

$$E(CS_i) = \int_{P_0}^{P_c} \lambda_i dP = -\frac{\lambda_i}{\beta_p}\Big|_{P_0}^{P_c} \tag{11}$$

where λ_i is as defined in (10) and β_p is the estimated parameter on travel cost. Summed across all i, the area measures the total per trip net willingness-to-pay by all individuals to recreate at the site. In the case of the ZINB model expected consumers surplus must be weighted by the probability of zero visitation ($1 - P_i$), where P_i is a function of variables that affect the participation decision:

$$E(CS_i) = (1 - P_i)\int_{P_0}^{P_c} \lambda_i dP = -(1 - P_i)\frac{\lambda_i}{\beta_p}\Big|_{P_0}^{P_c} \tag{12}$$

Application to Lake Sevan, Armenia

Lake Sevan is the largest high altitude reservoir of fresh water in the Transcaucasus, and is one of the highest lakes in the world. However, over the course of last 50 years, the level of the lake has dropped by 18 m, its surface area has decreased by 15%, and the volume of water in Lake Sevan fell by more than 40% (from 58.5 to 34.6 km^3). A significant proportion of this decrease occurred in the early 1930s when engineers attempted to drain the lake for agricultural purposes. However, it was found that the soils underlying the lake were unsuitable for cultivation. A few decades later agricultural expansion in the Ararat valley demanded large irrigation schemes of which Lake Sevan was the primary source of water. The level of water withdrawal was much larger than the recharge rates of inflowing rivers to Lake Sevan. The drop in the level of the lake had various significant adverse impacts on Lake Sevan's ecology. Most notable were the ecological impacts on endemic species of fish, birds and plants. Three unique species of trout went extinct and nesting birds fell in population. From a recreational standpoint, located only 70 km away from the capital city Yerevan, Lake Sevan remains the preferred and most accessible recreational site to most Armenians. However, the recreational experience has declined as the water level and quality fell.

In response to these perturbations, the Government of Armenia has been working on a Lake Sevan protection action plan. The objectives under consideration by the

Government of Armenia include preventing a further lowering of the level of Lake Sevan, and raising the level of the lake by at least 3 meters as quickly as possible. However to date, there has not been a thorough measurement of the current recreational benefits to include in benefit-cost analysis. Welfare measurement would be useful to policymakers tasked with weighing the alternative options of restoring Lake Sevan. Our model and welfare comparison is also useful in this context as Lake Sevan is a single site, with no substitutes, so comparing the two samples is not confounded by alternative sites that may enter into an individual's water-based recreation decision. Also, since we are measuring current recreational benefits, we avoid having to predict what impact the improvements would have on expected trip demand.

To estimate benefits by the general population and users of the site, two surveys were conducted – one comprising of 3,358 households across Armenia and the other an interceptor survey of 389 on-site tourists recreating at Lake Sevan. Both were conducted in the year 2000, with the tourist survey during the summer to better capture the high season of annual recreational use at the lake. The household sample was selected and stratified by clusters according to the latest Population Census of Armenia (with addresses) available at the time, in 1996. We also over-sampled in the capital city, Yerevan, because of the relative proximity of the higher income population to the site and also in the Lake Sevan area because of the relative dependence of people's daily lives on the lake. In-person interviews were conducted for the household survey, according to the stratified list of addresses from the Census. When no one was home, or were underage, the interviewer visited the next address from the stratified list. This process continued until a pre-specified minimum number of households in the cluster were met. The interviews took place over a three month period, and with a response rate of approximately 80% and a refusal rate of 10%.

The on-site survey consisted of in-person interviews as well among the beach, camping, day-use and cultural sites. Each of these areas was divided into "territorial clusters" such that a pre-specified number of interviews were eventually completed. Approximately 82% of the 389 respondents were from Yerevan city and since the tourist survey occurred before the household survey, respondents in the household sample were asked if they had been interviewed at the lake to avoid any overlap. The sampling took approximately one week to complete, with a final refusal rate of 3.5%.

From the empirical demand relationship in equation (10), we model the participation and trip quantity decisions using travel cost and several individual-specific variables that may co-vary with each decision - income, age, household size, education, and a Yerevan city dummy. Travel costs included: (1) transport costs; (2) on-site costs (per day); and (3) the value of time traveling *to* and spent at Lake Sevan. Respondents were asked what they spent on transport costs (fuel, vehicle rental costs, or transport ticket costs in the case of a tour), and on-site costs (food, beverages, lodging, entrance fees, parking fees, fishing, beach, picnic, boating, cultural site fees, plus any others that they may have stated). Transport costs and on-site costs were normalized to per person costs, and on-site costs were further normalized to per day costs in the case of a multi-day trip.

Respondents also were asked their opportunity cost of time spent traveling and on-site; i.e., the amount they could have earned during the entire trip had they not taken the trip. The value of time was calculated as the sum of the stated opportunity cost from the respondent and the time it took them to travel to the site (also given by the respondent). This value was then normalized to a per day value of time in the case of a multi-day trip. Previous research has shown that assumptions on time values are a primary determinant of the estimated values of recreation activities (McConnell and Strand, 1981; Bishop and Heberlein, 1979; Wilman and Pauls, 1987). Site values may vary four-fold, depending on the value of time (Fletcher *et al.*, 1990). However, there is a general lack of consensus of how to exactly measure this in survey work. Many of the suggested alternatives gravitate around some constant proportion of the wage rate (Bockstael *et al.*, 1987; McConnell and Strand, 1981). In Armenia, the simple use of this proxy may prove to be difficult given the extent of informal labor markets.[10] This led us to include informal sector earnings in defining total income as well as including a question of how much money the respondent would have earned if they had not taken the trip to Lake Sevan.

The context in which this question was asked is important since, at the time, Armenia faced tremendous unemployment as a consequence of the economic downturn since the fall of the Soviet regime; hence formal employment opportunities were scarce. A large proportion of a respondent's stated opportunity cost was from sporadic informal employment or entrepreneurial income. Thus we decided to take the full value of this amount rather than the conventional method in travel cost studies where it is some proportion of the wage. In reality, wage (or formal) income constituted only a nominal fraction of total income, and would not be an accurate measure of the tradeoff between labor and leisure decisions. We investigate whether the value of time influences the participation decision by including it in an alternative specification of the models.

Annual visitation to Lake Sevan by these two groups is reported in Table 1, with descriptive statistics in Appendix I. Household survey responses indicate that nearly 75% did not visit the lake in the past year, with a sample mean of 0.81 day-trips. The tourist survey, obviously truncated at one as interviews took place at the lake, averaged 3.17 day-trips per year. Travel costs (transport and on-site costs) reported by respondents averaged $7-8 USD per day, while the stated opportunity cost was surprisingly low at only $1.26-1.30 USD per day by those in the household and tourist survey, respectively. The average person from the household survey was 44 years old, earned the equivalent of $1,383 USD per annum, had 10 years of formal education, and a household size of 4. The average person from the on-site survey was 36 years old, earned $2,933 USD per annum, had 10 years of education and a household size of 5. The rather large difference in mean income between the two groups may be indicative of only those who can afford to recreate being on-site. Even though 82% of the on-site respondents stated they were from Yerevan, these may be folks who have higher-than-average income and thus have a higher likelihood of visitation than those among the general population of Yerevan.

In Table 1 we also note that the variance of visitation in each sample exceeds its mean, thus we suspect the presence of overdispersion, and therefore formally test the negative binomial counterpart of the Poisson distribution. In addition, given the large

[10] In fact, it is common knowledge that many Armenians derive a significant proportion of their income through transfers from relatives abroad rather than through their own labor (both formally and informally).

number of zeros in the household survey, we formally test the use of the zero-inflated negative binomial model for the household survey.

TABLE 1
Frequency of Visitation

Person-day-trips	Household Frequency	Percent	Tourist Frequency	Percent
0	2516	74.93	0	0.00
1	455	13.55	185	47.56
2	152	4.53	94	24.16
3	84	2.50	41	10.54
4	30	0.89	25	6.43
5	37	1.10	14	3.60
6	12	0.36	5	1.29
7	7	0.21	0	0.00
8	5	0.15	0	0.00
9	0	0.00	0	0.00
10	26	0.77	5	1.29
10 to 15	12	0.36	6	1.54
15 to 20	10	0.30	6	1.54
20 to 30	3	0.09	4	1.03
30 to 40	3	0.09	2	0.51
40 to 50	1	0.03	2	0.51
50 to 100	5	0.15	0	0.00
Total	3358	100.00	389	100.00
Mean	0.81		3.17	
Standard deviation	3.95		5.75	

Estimation Results

(i) Determinants of visitation

The household sample was initially modeled using the Poisson, negative binomial (NB), zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB). The on-site sample was modeled using the truncated Poisson, truncated negative binomial (TRNB) and the truncated negative binomial with endogenous stratification (TRNBES). Comparative tests between each model were performed and are reported below. For brevity, only the estimation results for the household NB and ZINB and on-site models TRNB and TRNBES are reported in Table 2 with marginal effects for the ZINB and TRNBES models listed in Table 3. Note that for the household model, each equation (*logit inflation model* and *negative binomial model*) contains the same explanatory variables as they may contribute to either the participation or frequency of visitation decisions.

We also present two specifications for each of the econometric models above, one with travel costs in aggregate and another separating out the value of time. Beginning with the household survey results in the second through fifth columns of Table 2, we

Appendix 1: Descriptive statistics for the Household (HH) and Tourist survey (Tourist)

Variable	*Mean*			*Standard deviation*			*Minimum*			*Maximum*		
	HH w/ Trips ≥ 0	*HH w/ Trips > 0*	*Tourist Trips ≥ 1*	*HH w/ Trips ≥ 0*	*HH w/ Trips > 0*	*Tourist Trips ≥ 1*	*HH w/ Trips ≥ 0*	*HH w/ Trips > 0*	*Tourist Trips ≥ 1*	*HH w/ Trips ≥ 0*	*HH w/ Trips > 0*	*Tourist Trips ≥ 1*
Visits (person-day-trips)	0.81	3.24	3.17	3.95	7.36	5.75	0	1	1	100	100	50
Travel costs ($USD)[1]	7.74	8.14	8.94	4.39	8.76	6.39	0.06	0.06	0.1	113	113	41
Value of time ($USD)	1.26	1.26	1.30	1.84	3.67	3.34	0	0	0	50	50	21
Income ($USD)	1,383	1,861	2,933	1,246	1,623	2,052	120	150	480	14,976	14,976	15,120
Age (years)	44	39	36	14	12	13	18	18	18	81	76	71
Household size	4	5	5	2	2	1	1	1	2	13	12	8
Education (years)	10	11	10	2	2	2	0	0	5	14	14	14
Past visitation (1=yes)	0.95	1.0	0.94	0.22	0	0.24	0	1	0	1	1	1
Yerevan city (1=yes)	0.82	0.80	-	0.38	0.40	-	0	0	-	1	1	-
Lake Sevan (1=yes)	0.06	0.12	1.00	0.24	0.33	0.00	0	0	1	1	1	1
Observations	3358	842	389									

[1]Includes transport and on-site costs; excludes the value of time.

note that the likelihood ratio (LR) test of $\alpha = 0$ is rejected indicating the significant presence of overdispersion and thus we select the negative binomial specification over the Poisson. A formal specification test between the NB and ZINB was conducted (Vuong, 1989). The test statistic is directional and distributed standard normal; for values $|V| > 1.96$, the zero-inflated version is supported. With values of 4.86 and 4.79, the two ZINB specifications are favored over the two NB.

Parameter estimates of the household ZINB model reveal that income, age and education, along with respondents who reside in Yerevan significantly determine the household participation decision to recreate at Lake Sevan (see *logit inflation model*). Note that the coefficients are interpreted relative to observing a zero count. Hence, the positive coefficient on age implies, perhaps as expected, that older respondents are more likely to record zero participation. We also find that individuals with higher income or higher level of education are more likely to report a positive number of trips (participate) to Lake Sevan. Finally, with respect to the participation decision, we find that Yerevan city residents are more likely to report zero visitations in the past year. It is not immediately clear why this may be so. However, we note that the mean per capita income of Yerevan residents is less than on-site participants' mean income which may thus constrain their participation decision. Even within the on-site sample, the result implies people in Yerevan are less likely to visit than those outside Yerevan. As pointed out by a reviewer, it is also of interest to note a significant difference in car ownership. In the household survey, 28% of individuals owned a car, while 63% of the individuals in the on-site survey owned a car. It is thus possible that accessibility could be an issue for those in Yerevan, explaining the lower probability of visitation.

Among those who do choose to participate (see *negative binomial model*), we find that increases in income and household size increase trip demand. Increases in income thus appear to increase both participation and visitation to Lake Sevan. However, increases in education decrease trip demand (while increasing participation). Finally, we find (perhaps as expected) that increases in travel costs decrease trip demand. Note however that the decision to participate is not influenced by travel costs (or the opportunity cost of time). Although this may not be a general finding, Gurmu and Trivedi (1996) also find the participation decision to be insensitive to travel costs.

For the on-site survey, the LR test between a truncated Poisson and truncated negative binomial (TRNB) was rejected indicating that overdispersion in visitation is significant, leading us to favor the TRNB specification. The TRNBES model was also estimated to see whether higher trip frequencies have any systematic association with an individual's characteristics. Estimation results for both TRNB and TRNBES show that increases in travel costs, age and education decrease visitation, whereas increases in household size increase trip demand. These results are consistent to those found in the household survey. However, contrary to the household model, we find the value of time to be significant in the decision to recreate by those on-site.

From a statistical standpoint the estimated coefficients and standard errors are higher in the TRNB model leading to a lower significance across each explanatory variable. This is contrary to theory which says that before accounting for endogenous stratification the standard errors are smaller, leading to inflated significance in the estimates (Loomis, 2003; Martinez-Espineira and Amoako-Tuffour, 2005). In our case, once we correct for endogenous stratification, the magnitude of the estimated coefficients falls, and standard errors fall by a greater extent such that the statistical significance rises among the major determinants of visitation. In the next section, we explore the consequences of these differences on expected trip demand as well as the implications on welfare estimates.

(ii) Visitation sensitivity

The sensitivity of trip demand for the household ZINB and tourist TRNBES models to changes in the independent variables is summarized in Table 3. Note that the two models presented are those that had the value of time as a separate regressor. The estimated coefficients in the *logit inflation equation* for the ZINB model are interpreted as the impact on the log-odds of non-participation (i.e., the log-odds of observing a zero), which may be somewhat confusing; therefore, we reverse the signs on the estimated coefficients and directly interpret the results in terms of the log-odds of *participation*.

A unitary increase in the age of the respondent leads to a decrease in likelihood of participation by 9.4%, whereas an increase in one year of education increases the log-odds of participation by 24%. Travel costs, the value of time and household size are insignificant in estimation. Income only marginally impacts trip demand; increase of $1 USD leads to a 0.12% increase in participation. Residing in Yerevan has a large, negative impact on participation of 135%. For the trip count equation, a one unit increase in travel costs or education decreases the number of trips by 2.3% and 6.3%, respectively. Thus, even though travel costs are not a significant determinant in the trip participation decision, they are highly significant to the number of trips a person decides to take; albeit the impact is small (2.3%). Also, as noted previously, a person's education level appears to be important to both decisions, but in opposite directions. Those with one additional year of education are 24% more likely to participate and for those who already participate, one additional year actually decreases the number of trips taken by 6.3%. Household size was found to be insignificant in the participation equation, but for those who do recreate, a one unit change in household size increases the number of trips by 9.4%. Upon closer inspection of the data, it was found that households with more children were associated with higher trip frequencies. Income was found to be statistically significant albeit with a negligible impact on trip frequency.

For on-site trip demand, unitary increases in travel costs, the value of time, age and education decrease the number of trips by 4.3%, 7.2%, 2.5% and 9.4%, respectively, and an increase in household size significantly increases trip frequency by 33%. With the exception of age, each impact has a similar interpretation as in the household model, but the effects are much larger. In the case of age, older individuals are significantly and negatively correlated with higher visitation.

TABLE 2: Household and on-site model estimates of visitation to Lake Sevan

Variable	*Household: NB1*	*Household: NB2*	*Household: ZINB1*	*Household: ZINB2*	*On-site: TRNB1*	*On-site: TRNB2*	*On-site: TRNBES1*	*On-site: TRNBES2*
Negative binomial model								
Travel costs (with VOT†)	-0.0262***		-0.0157***		-0.0504***		-0.0506***	
	(-5.57)		(-3.53)		(-3.31)		(-4.71)	
Travel costs (without VOT†)		-0.0351***		-0.0231***		-0.0421**		-0.0438***
		(-5.96)		(-4.06)		(-2.42)		(-3.50)
Value of time		0.0013		0.0073		-0.0802**		-0.0749***
		(0.10)		(0.57)		(-2.41)		(-2.84)
Income	0.00036***	0.00035** *	0.00015***	0.00015***	0.000040	0.000034	0.000011	0.000008
	(7.54)	(7.62)	(3.67)	(3.68)	(0.55)	(0.51)	(0.27)	(0.22)
Age	-0.0234***	-0.0238***	0.0034	0.0030	-0.0307***	-0.0299***	-0.0257***	-0.0252***
	(-6.39)	(-6.52)	(0.76)	(0.67)	(-3.38)	(-3.28)	(-4.48)	(-4.39)
Household size	0.1224***	0.1163***	0.0980***	0.0900***	0.2993***	0.3207***	0.2734***	0.2850***
	(4.04)	(3.85)	(2.66)	(2.45)	(3.59)	(3.73)	(5.31)	(5.40)
Education	-0.0098	-0.0061	-0.0690***	-0.0649***	-0.0967*	-0.0954*	-0.0972***	-0.0982***
	(-0.45)	(-0.28)	(-2.68)	(-2.52)	(-1.78)	(-1.75)	(-2.95)	(-2.98)
Constant	-0.0318	0.0022	0.2219	0.2538	-9.5993	-8.9529	-16.9634	-15.4655
	(-0.09)	(0.01)	(0.57)	(0.66)	(-0.37)	(-0.33)	(-0.09)	(-0.07)
Logit inflation model								
Travel costs (with VOT†)			0.0121					
			(1.00)					
Travel costs (without VOT†)				0.0121				
				(0.87)				
Value of time				0.0112				
				(0.40)				
Income			-0.0012***	-0.0012***				
			(-4.75)	(-4.72)				
Age			0.0903***	0.0900***				
			(8.46)	(8.41)				
Household size			0.0318	0.0246				
			(0.44)	(0.34)				
Education			-0.2775***	-0.2748***				
			(-4.80)	(-4.75)				

TABLE 2: (CONTNUED)

Yerevan city			0.8574***	0.8568***				
			(2.65)	(2.64)				
Constant			-1.5628*	-1.5421*				
			(-1.83)	(-1.81)				
α	5.7930	5.7524	3.7112	3.6927	12.1417	11.3223	18.4937	16.9087
Log-likelihood	-3,334.03	-3,331.09	-3,249.31	-3,247.23	-656.73	-656.24	-680.26	-679.73
LR test (α=0) ~ χ^2 (d.f.)	6,463.91 (1)	6,301.00 (1)	3,274.44 (1)	3,236.80 (1)	848.18 (1)	849.09 (1)	801.12 (1)	802.10 (1)
Vuong test ~ N (0,1)	–	–	4.86	4.79	–	–	–	–
Number of observations	3,358	3,358	3,358	3,358	389	389	389	389
Non-zero observations	842	842	842	842	389	389	389	389
Zero observations	2,516	2,516	2,516	2,516				

t-statistics in parentheses; * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

†VOT = value of time.

TABLE 3: Sensitivity Analysis of Trip Demand

	Household: ZINB2		On-site: TRNBES2	
Visits	*Coefficient*	*% Δ trips*	*Coefficient*	*% Δ trips*
Negative binomial model				
Travel costs ($USD)	-0.0231***	-2.29	-0.0438***	-4.29
Value of time ($USD)	0.0073	0.74	-0.0749***	-7.22
Income ($USD)	0.00015***	0.02	0.000008	0.00
Age (years)	0.0030	0.30	-0.0252***	-2.49
Household size (number)	0.0900***	9.42	0.2850***	32.97
Education (years)	-0.0649***	-6.28	-0.0982***	-9.35
Pr(*Participation*)		% Δ Pr(*participation*)		
Logit inflation model				
Travel costs ($USD)	-0.0121	-1.22		
Value of time ($USD)	-0.0112	-1.13		
Income ($USD)	0.0012***	0.12		
Age (years)	-0.0900***	-9.42		
Household size (number)	-0.0246	-2.49		
Education (years)	0.2748***	24.03		
Yerevan (1=lives in Yerevan)	-0.8568***	-135.36		

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level

(iii) Estimated trip demand and welfare measures

Using the parameter estimates from the four models in Table 2, the expected number of trips, $E(y_i \mid \bar{X})$, and consumers surplus (CS) measures were calculated (Table 4). The expected number of trips was estimated for each model using sample means of the independent variables. Comparing the NB with the ZINB, note that the expected number of trips falls once we account for the inflation of zeros (participation). Indeed, since the NB model is treating every zero as being a part of the quantity decision, this biases the estimates upwards, whereas the ZINB recognizes that the zeros may come from different stochastic processes (participation or frequency of visitation).

For the on-site model, TRNB, the expected number of trips far exceeds the demand estimated by the household survey. This seems reasonable since we are comparing casual versus avid users of the site. However, the expected number of trips is even higher after accounting for ES (TRNBES). At first glance this may seem counter-intuitive, but recall that expected trip demand is calculated as $E(y_i \mid x_i) = \lambda_i + 1 + \alpha\lambda_i$), and note that the only substantial difference between the estimated parameters of TRNB and TRNBES is the value of the overdispersion parameter, α (see Table 2). Thus it is the overdispersion that is driving this result.

In Table 4, estimated household consumers surplus per trip was $8.33 for the NB and $8.40 for the ZINB model whereas for the on-site sample CS was calculated as

\$8.73 without compensating for endogenous stratification and \$8.21 per trip with endogenous stratification. This agrees with our prior expectations where after accounting for excess zeros in the household model, and truncation and endogenous stratification in the on-site model, the welfare measures are similar. A difference of means test reveals that the values are statistically different (at the 1% level of significance) between the household NB and ZINB, and between the on-site TRNB and TRNBES models. However, the estimated values are not statistically different between the household NB and on-site TRNB, and between the household ZINB and the on-site TRNBES model. This finding is similar to other studies where the bias of endogenous stratification in welfare measurement has been found to be significant (e.g Ovaskainen *et al.* 2001; Loomis, 2003; Martinez-Espineira and Amoako-Tuffour, 2005). The direction of bias is also of interest. In this study, the correction led to a smaller consumers surplus, as was the case in Loomis (2003) and Martinez-Espineira and Amoako-Tuffour (2005). One explanation could be that uncorrected estimates are inflated upwards owing to the characteristics of avid-natured tourists. If this segment of the population represents a significant proportion of the on-site sample, and without endogenous stratification correction, one may expect inflated welfare estimates. In our sample we did not find a large deviation possibly since characteristics of on-site individuals being similar to those in the household sample. The direction of bias is largely dictated by the composition of avid tourists in the sample, and the characteristic differences between them and the household population. Further research into the direction of this correction-bias may serve to establish this result in a more concrete manner.

TABLE 4: Expected visitation and benefit estimates

Measure	*Household: NB2*	*Household: ZINB2*	*On-site: TRNB2*	*On-site: TRNBES2*
$E(y_i \mid \bar{X})$	0.8838	0.5667	5.9251	7.4072
CS (\$USD per day-trip)	8.33 (3.38) (0.344)	8.40 (3.06) (0.327)	8.73 (3.55) (0.360)	8.21 (2.99) (0.320)
Total WTP[1] (\$USD)	6,488,226	6,549,617	6,802,126	6,399,840

Note: $\overline{X}$ is evaluated at the sample mean. Standard deviations are in the top parenthesis, standard errors in the lower.
1 – Calculated for households as: CS * 779,230 total households in Armenia according to 2001 Census.

Conclusion

In this paper, we have modeled a population-based household sample and an on-site sample in a single-site travel cost framework to compare estimated consumers surplus for the value of site access. In the household model, we have accounted for the potential for overdispersion by using a negative binomial distribution function, and for the possibility of observing a large number of zero visits (a recreation participation decision) by splitting the participation and frequency of visitation decisions directly in one censored model, the zero-inflated negative binomial. For the on-site survey, the distribution function is truncated at one. In addition, given the survey is on-site, there is a possibility of over-sampling those who recreate quite often; therefore the truncated distribution function is further augmented for endogenous stratification. To compare the effect of endogenous stratification, we have modeled the on-site sample as a truncated negative binomial with and without endogenous stratification.

After accounting for excess zeros in the household model, as well as for truncation and endogenous stratification in the on-site model, the welfare measures were not found to be statistically different. We find these results to be of great interest as researchers often find themselves, for reasons of resource and/or time limitations, with having to select and implement only one sampling methodology, thus leaving themselves vulnerable to one of two criticism: (1) if on-site survey is selected, then the sampling methodology is not representative of the entire population; or (2) if the household survey is selected, then the sampling methodology is not sufficiently representative of those actually recreating at the lake. Our results show that either method can be used to derive a consistent welfare measure of access to the site after accounting for specific statistical issues encountered in both approaches. An interesting avenue of future research would be to see whether these findings hold in the case of multi-site travel cost models.

References

Bishop, R.C. and T.A. Heberlein (1979). Measuring values of extramarket goods: are indirect measures biased? *American Journal of Agricultural Economics*, *61*(5), 926-932.

Bockstael, N., I. Strand, and M. Hannemann (1987). Time and the recreational demand model. *American Journal of Agricultural Economics*, *69*(2), 293-302.

Cameron, A. C. and P. K. Trivedi (1986). Econometric models based on count data: comparisons and application of some estimators and tests. *Journal of Applied Econometrics*, *1*, 29-53.

Cameron, A. C. and P. K. Trivedi (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, *46*(3), 347–364.

Cameron, A. C. and P. K. Trivedi (2001). Essentials of count data regression, in B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, pp. 331–348, Blackwell, Oxford, U.K.

Creel, M. and J. B. Loomis (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics, 72*, 434–441.

Curtis, J. (2003). Demand for water-based leisure activity. *Journal of Environmental Planning and Management*, *46*(1), 65-77.

Englin, J. and J. S. Shonkwiler (1995). Estimating social welfare using count data models: an application to long-run recreational demand under conditions of endogenous stratification and truncation. *Review of Economics & Statistics*, *77*(1), 104-112.

Englin, J. T. Holmes and E. Sills (2003). Estimating forest recreation demand using count data models, in E. Sills (ed.), *Forests in a Market Economy*, Chapter 19, pp. 341-359, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Fletcher, J. J., W. L. Adamowicz, T. Graham-Tomasi. (1990). The travel cost model of recreation demand: theoretical and empirical issues. *Leisure Sciences*, *12*, 119-147.

Gourieroux, C. A., A. Monfort, A. Trogon (1984). Pseudo maximum likelihood methods: applications. *Econometrica*, *52*, 701-720.

Greene, W. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working Paper EC-94-10, Department of Economics, Stern School of Business, New York University, New York, N.Y.

Grogger, J. and R. Carson (1991). Models for truncated counts. *Journal of Applied Economics*, *6*, 225-238.

Gurmu, S. and P. K. Trivedi (1996). Excess zeros in count models for recreational trips. *Journal of Business and Economics Statistics, 14*, 469-477.

Haab, T. C. and K. E. McConnell (1996). Count data models and the problem of zeros in recreation demand analysis. *American Journal of Agricultural Economics*, *78*, 89-102.

Hausman, J., B. Hall, Z. Griliches (1984). Econometric models for count data with an application to the patents – R&D relationship. *Econometrica*, *52*, 909-938.

Hellerstein, D. M. (1991). Using count data models in travel cost analysis with aggregate data. *American Journal of Agricultural Economics*, *73*, 860-866.

Lambert, D. (1992). Zero-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*, 1-14.

Loomis, J. (2003). Travel cost demand model based river recreation benefit estimates with on-site and household surveys: comparative results and a correction procedure. *Water Resources Research*, *39*(4), 1105.

Martinez-Espineira, R. and J. Amoako-Tuffour (2005). Recreation demand analysis under truncation, overdispersion, and endogenous stratification: an application to Gros Morne National Park. Working Paper 2005-03. Department of Economics, St. Francis Xavier University: Canada.

McConnell, K. and I. Strand (1981). Measuring the cost of time in recreation demand analysis: an application to sportfishing. *American Journal of Agricultural Economics*, *63*(Feb.), 169-173.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*, 341-365.

Ovaskainen, K., J. Mikkola and E. Pouta (2001). Estimating recreation demand with on-site data: an application of truncated and endogenously stratified count data models. *Journal of Forest Economics*, *7*(2), 125-144.

Parsons, G. R. (2003). The travel cost model. In P.A. Champ, K.J. Boyle and T.C. Brown (eds.), *A Primer on Nonmarket Valuation*. Kluwer Academic Publishers. Pp. 269-329.

Scrogin, D., K. Boyle, G. Parsons and A. Plantinga (2004). Effects of regulations on expected catch, expected harvest, and site choice of recreational anglers. *American Journal of Agricultural Economics*, *86*(4), 963-974.

Shaw, D. (1988). On-site samples' regression: problems of non-negative integers, truncation and endogenous stratification, *Journal of Econometrics*, *37*, 211-223.

Shaw, W. D. and P. Jakus (1996). Travel cost models of the demand for rock climbing, *Agriculture and Resource Economics Review*, *25*, 133-142.

Shaw, W. D., E. Fadali, and F. Lupi (2003). Comparing consumer's surplus estimates calculated from intercept and general survey data. Proceedings of the W-133 (U.S.D.A.) Regional Economics Group, compiled by J. S. Shonkwiler. Las Vegas, Nevada, February.

Shonkwiler, J. S. and W. D. Shaw (1996). Hurdle count-data models in recreation demand analysis, *Journal of Agricultural and Resource Economics*, *21*, 210-219.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, *57*, 307-334.

Wilman, E.A. and R.J. Pauls (1987). Sensitivity of consumers' surplus estimates to variation in the parameters of the travel cost model, *Canadian Journal of Agricultural Economics*, *35*, 197-211.

Yen, S. T. and W. L. Adamowicz (1993). Statistical properties of welfare measures from count data models of recreation demand. *Review of Agricultural Economics, 15*, 203–215.