

RESEARCH ARTICLE

Effects of *Ad-hoc* Data Truncation and Homogeneous Preferences on Recreational Demand and Values: An Application to the George Washington and Jefferson National Forests

Kavita Sardana^{1*} , John C. Bergstrom² and J. M. Bowker³

¹TERI School of Advanced Studies, 10 Institutional Area, Vasant Kunj, New Delhi, India, ²Department of Agricultural and Applied Economics, The University of Georgia, Conner Hall, Athens, GA, USA and ³Southern Research Station, U.S.D.A. Forest Service, USA

*Corresponding author. Email: kavita.sardana@terisas.ac.in

Abstract

We estimate a travel cost model for the George Washington & Jefferson National Forests using an On-Site Latent Class Poisson Model. We show that the constraints of *ad-hoc* truncation and homogenous preferences significantly impact consumer surplus estimates derived from the on-site travel cost model. By relaxing the constraints, we show that more than one class of visitors with unique preferences exists in the population. The resulting demand functions, price responsive behaviors, and consumer surplus estimates reflect differences across these classes of visitors. With heterogeneous preferences, a group of 'local residents' exists with a probability of 8% and, on average take 113 visits.

Keywords: latent class models; on-site Poisson model; recreational demand models; *ad-hoc* truncation; homogeneous preferences

JEL Classifications: Q51; C14

1. Introduction

In previous studies of recreation demand, researchers have dropped visitor observations with very high-frequency trips to recreational sites who may be considered outliers. For example, Englin and Shonkwiler (1995) drop observations with annual trips greater than 12, allowing one trip per month. Egan and Herriges (2006), Bowker et al. (2009), and Sardana, Bergstrom, and Bowker (2016) drop observations with annual trips greater than 52, allowing one trip per weekend. By dropping these observations, researchers truncate the distribution of visitors to the site, and the truncation is somewhat *ad-hoc*, that is, the truncation rule or threshold does not originate from theory.

These visitors with higher annual trips may represent local residents who report many trips due to close proximity of the recreation site to their place of residence (e.g., home). These local residents who take frequent short-duration trips may, for instance, engage in activities as a regular part of their exercise routine (e.g., jogging, taking a walk) and incur a lower travel cost. They derive higher consumer surplus per visit and due to more frequent visits, their aggregate consumer surplus per individual is un-proportionally higher.

However, to create one homogeneous class of visitors, previous recreation demand studies including those mentioned above model the visitor population as one recreation visitor group with homogeneous preferences. The distribution of trips taken to recreational site by this

homogenous group is left-side truncated at zero, and perhaps also right-side truncated by the researcher at some rather *ad-hoc* cut-off or threshold number of trips. There are at least two issues involved with modeling recreation demand imposing homogeneous visitor preferences and *ad-hoc*, right-side truncation to the distribution of trips. First, *ad-hoc* truncation imposes a constraint on the estimated model which is likely different from the actual, underlying conceptual model. Second, homogenous preferences impose another constraint on the estimated model which could result in incorrect estimation of consumer surplus. The latter issue has been treated by Parsons (1991) using an instrumental variable approach and Baerenklau (2010) using a latent variable approach.

Ad-hoc truncation can be addressed by allowing the visitor population to be heterogeneous. Baerenklau (2010) estimates a zonal travel cost model with heterogeneous preferences using a Poisson specification for backcountry hikers for a Southern California study site. They found that the welfare estimates under the constraint of homogeneous preferences were substantially higher as compared to the unconstrained model. Hynes and Greene (2013) combined revealed and contingent travel data to model heterogeneous preferences using an on-site Negative Binomial specification for beach visitors in Ireland. Martinez-Cruz and Sainz-Santamaria (2015) estimated an on-site travel cost model with heterogeneous preferences for peri-urban forests in Mexico City. In both of these papers, it was found that three classes of visitors exist with unique preferences and welfare estimates.

In this paper, we show how *ad-hoc* truncation constrains the estimated model under homogeneous preferences and how these constraints can be handled using a latent variable approach which allows for heterogeneous preferences. Whether or not the assumption of homogeneous preferences leads to underestimation or overestimation of welfare effects (e.g., consumer surplus) is an empirical question that we seek to answer in the paper using the latent variable approach for modeling recreational trips to the George Washington and Jefferson National Forests in the U.S. southeastern region.

2. Theoretical Considerations

2.1. Ad-Hoc Truncation, Homogenous Preferences, and Heterogeneous Preferences

As discussed above, we argue that high-frequency visitors may not be engaging in day visits to the site in question as conventionally thought of in recreational demand modeling, but rather are engaging simply in, say, a daily jog or walk which is part of a broader activity, such as a “daily exercise routine”. This broader “daily exercise routine” is not necessarily dependent on developed recreational facilities. Dropping these individuals or modeling these individuals using a recreation demand framework with homogeneous preferences would be incorrect from a theoretical perspective as explained below. In addition, classifying visitors into a homogeneous class of visitors who mostly engage in developed activities is incorrect from a policy perspective because resource managers who manage decisions control multiple recreational settings such as developed, wilderness, and general undeveloped forest areas¹ (e.g., see Sardana, Bergstrom, and Bowker, 2016).

¹The National Visitor Use Monitoring Program (NVUM) of the U.S. Department of Agriculture (USDA) Forest Service classifies settings into four categories: Wilderness (WILD), Overnight-use Developed Settings (OUDS), Day-use Developed Settings (DUDS), and General Forest Areas (GFA). WILD areas are officially designated wilderness subject to the provisions of the U.S. Wilderness Act of 1964. DUDS have facilities for day-use activities including picnicking, boating, and developed-trail hiking. OUDS have facilities for overnight stays for activities such as developed camping. GFA are areas which have undeveloped facilities for activities like nature viewing, hunting, developed and undeveloped trail, hiking, and some motorized sports (English et al., 2002). According to the NVUM FY 2012-FY 2016 National Summary Report, visits to Wilderness and General Forest Area settings throughout the National Forest System were estimated at 99,544,000 visits for the period FY 2012-FY 2016, which is 67% of total National Forest visits. <https://www.fs.fed.us/recreation/programs/nvum/pdf/5082016NationalSummaryReport062217.pdf>

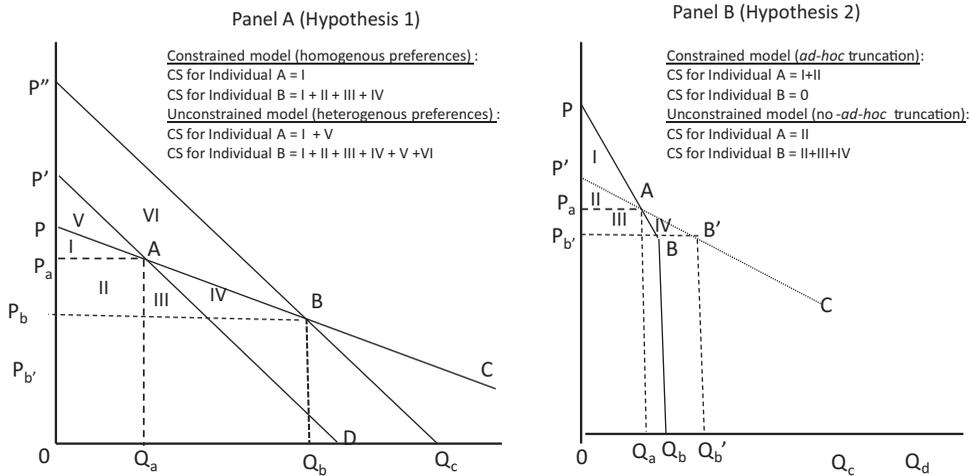


Figure 1. Consumer surplus (CS) under constrained and unconstrained models (assumptions).

The effects of homogenous preferences, heterogeneous preferences, and *ad-hoc* truncation on consumer surplus estimates are illustrated in Figure 1. First, consider the effects of homogenous preferences vs. heterogeneous preferences. In Figure 1, Panel A, the demand curve under homogenous preferences is given by PC for both Consumers A and B.

Given the demand curve under homogenous preferences in Panel A, Consumer A consumes Q_a trips at a price of P_a with consumer surplus equal to area I, and Consumer B consumes Q_b trips at a price of P_b with consumer surplus equal to area I + II + III + IV.

In Panel A, under the assumption of heterogeneous preferences, the demand curve for Consumer A is given by P'D and the demand curve for Consumer B is given by P''Q_c. Given these demand curves, Consumer A again consumes Q_a trips at a price of P_a with consumer surplus equal to area I + V. Consumer B consumes Q_b trips with consumer surplus equal to area I + II + III + IV + V + VI.

For Consumer A, the difference in consumer surplus under homogenous preferences and consumer surplus under heterogeneous preferences is equal to $(I) - (I + V) = -V$. For Consumer B, the difference in consumer surplus under homogenous preferences and consumer surplus under heterogeneous preferences is equal to $(I + II + III + IV) - (I + II + III + IV + V + VI) = -(V + VI)$. Thus, for both Consumer A and Consumer B, given the demand curves in Panel A, we expect that consumer surplus under heterogeneous preferences will exceed consumer surplus under homogenous preferences, leading to the following hypothesis:

Hypothesis 1: The assumption of homogenous preferences will lead to an underestimation of consumer surplus as compared to the assumption of heterogeneous preferences.

Panel B illustrates the consumer surplus per individual with *ad-hoc* truncation. The forms of truncation illustrated in Figure 1, in general, appear in previous studies using on-site survey samples when truncation limits are treated as exogenous or pre-determined. For on-site samples, the left-truncation limit is set at zero (Haab and McConnell, 2002, pp. 174–181). For right-truncation, the limit is somewhat *ad-hoc* and is dependent on a researcher's subjective definition of the cut-off point for the maximum number of trips for high-frequency visitors.

For example, in Figure 1, Panel B, the point of *ad-hoc* truncation is given by Point B. In Panel B, under *ad-hoc* truncation, the demand curve for both Consumers A and B is given by PBQ_b. Given

this demand curve, Consumer A consumes Q_a trips at a price of P_a with consumer surplus equal to area I + II and Consumer B who consumes Q_b trips has consumer surplus equal to 0, as consumer B's demand gets cut-off due to *ad-hoc* truncation.

In Panel B, the demand curve without *ad-hoc* truncation is given by $P'B'C$. Given this demand curve, Consumer A consumes Q_a trips at a price of P_a with consumer surplus equal to area II. Consumer B consumes Q_b trips at a price of P_b with consumer surplus equal to area II + III + IV. The difference in consumer surplus between the demand curve without *ad-hoc* truncation and the demand curve for this consumer assuming *ad-hoc* truncation equals to $(2II + III + IV) - (I + II) = (II + III + IV - I)$. Thus, imposing the constraint of *ad-hoc* truncation, the net change in consumer surplus depends on the relative size of area I as compared to area $(II + III + IV)$, leading to the following hypothesis:

Hypothesis 2: The assumption of *ad-hoc* truncation will overestimate/underestimate the consumer surplus depending on the relative size of consumer surplus in the unconstrained model for the individual whose demand gets cut-off due to imposition of the constraint of *ad-hoc* truncation.

2.2. Travel Cost Method Demand Modeling with Heterogeneous Preferences

In order to account for the more realistic assumption of heterogeneous visitor-type preferences, we employed the Latent Variable approach that identifies distinct groups of individuals, but each group is assumed to be taking visits solely for recreation. This model is a type of latent class model with distinct groups. Latent class models are in the class of finite mixture models where heterogeneous individual preferences are modeled as a mixture of distinct but unobservable groups (Wedel et al., 1993). Modeling heterogeneous preferences is essential if the researcher believes that in the population, there exist at least two types of people who are heterogeneous in the sense that their marginal effects are different. Assigning everyone *a priori* into one class or another based on household socio-economic demographics is incorrect since these household characteristics are imperfectly observed.

Single-site-based recreation trip demand and the value of recreation site access were estimated using the travel cost method (TCM). The TCM is a revealed preference nonmarket valuation technique that uses costs incurred by an individual or group traveling from their origin (e.g., primary residence) to the destination as a proxy for trip price. Price (travel cost) and quantity (number of trips) data can be used to estimate a demand function that is applied to estimate trip demand and welfare effects in the form of consumer surplus (Freeman, 1992).

A single-site TCM recreation demand function corresponds theoretically to the Marshallian demand function of the general form,

$$y^i = y^i * (p_y, M, q), \quad (1)$$

where the dependent variable in (1), y^i , represents annual trips to a recreation site by an individual or group i , p_y represents the full travel cost to an individual or group, M represents socio-economic demographics of an individual or group, and q represents quality of the site (which is constant in the single-site model). Because recreation trips by nature are non-negative integer values, the dependent variable in (1) takes on non-negative integer values. Thus, the ordinary least-squares regression model is inappropriate to estimate the demand model since the data-generating process for non-negative integer values is a count model. The basic model that satisfies the non-negative integer or the count data process is the Poisson model (Hellerstein, 1991).

Predominant problems with on-site samples are truncation, that is, exclusion of non-users (Borzykowski, Baranzini, and Maradan, 2017), and endogenous stratification, that is, oversampling frequent visitors. Shaw (1988) and Englin and Shonkwiler (1995) derived the distribution

correcting for the joint effects of truncation and endogenous stratification (on-site) for the Poisson and Negative Binomial distribution, respectively. The conditional probability density for the on-site Poisson model (Shaw, 1988) is given in (2),

$$Pi|s(y_i|\beta_s) = \frac{\exp(-\lambda_i)\lambda_i^{(y_i-1)}}{(y_i - 1)!} \tag{2}$$

The probability that an individual *i* belongs to group *s* ∈ {1, ..., S} is assumed to be logistic π_s (Baerenklau, 2010, p.804) as shown in (3). In (3), φ_s is an estimable parameter vector and z_i are individual characteristics that influence group membership,

$$\pi_s = \frac{\exp(\phi'_s z_i)}{\sum_{s=1}^S \exp(\phi'_s z_i)} \tag{3}$$

φ₁ = 0, for identification

The formulation of the probability density in (4) is conditional upon subject *i* belonging to class *s*. Considering the observed frequencies of *y_i* as arising from an unobserved, mixture Poisson distribution, one obtains the unconditional probability density,

$$Pi(y_i|\beta_s) = \sum_{s=1}^S \pi_s \frac{\exp(-\lambda_{i,s})\lambda_{i,s}^{(y_i-1)}}{(y_i - 1)!} \tag{4}$$

The likelihood function for estimating the parameters in (4) is given by,

$$L = \prod_{i=1}^n Pi(y_i|\beta_s) = \prod_{i=1}^n \sum_{s=1}^S \pi_s [-\lambda_{i,s} + (y_i - 1)(X_i\beta_s) - \ln[(y_i - 1)!] \tag{5}$$

Based on the estimated parameters, everyone in the population of interest can be assigned to one of the classes based on posterior probabilities. The unconditional posterior probability associated with *S* latent segments is given by (6):

$$\theta_{is}^* = \frac{\hat{\pi}_s \hat{p}_{i|s}(y_i|\beta_s)}{\sum_{s=1}^S \hat{\pi}_s \hat{p}_{i|s}(y_i|\beta_s)} \tag{6}$$

The maximum likelihood estimates of β_s can be obtained by maximizing the likelihood function in equation (5). Heckman and Singer (1984) advocated the use of an expectation maximization algorithm (Dempster, Laird, and Rubin, 1977) for this type of problem. For the Latent Variable model, the researcher specifies *S*, the number of segments. The number of *S* can be determined based on the consistent Akaike information criterion (CAIC) (Bozdogan, 1987) given by (7):

$$CAIC = -2 \sum_{s=1}^S LL_S + K(\ln(N) + 1) \tag{7}$$

where LL_S is the log-likelihood value, *K* is the number of parameters, and *N* is the number of observations.

Consumer surplus is estimated from the empirical demand model based on the following calculations. Following Shaw (1988), the conditional expected value of *y* given *x* (individual characteristics that influence demand) for the on-site Poisson model is given by (8):

$$E(y|x) = \exp(x\beta) + 1 \tag{8}$$

The conditional expected value of *y* given *x* for each class in the Latent Variable (latent class) model jointly corrected for truncation and endogenous stratification is therefore given by (9):

Table 1. Descriptive statistics

Variable	Description	N	Mean	Std. Dev.	Min	Max
Visits	Number of visits to the forest for the activity indicated as interviewee's primary activity	237	23.722	44.696	1	260
Income	Income category for total household income of respondent (thousand dollars)	237	59.458	31.863	25	124.5
Female	Gender of the respondent	237	0.359	0.481	0	1
Age	Age of the respondent	237	41.181	15.607	17.5	70
TC	Travel cost of the respondent (dollars)	237	14.939	26.629	0.124	185.853
PeopleVeh	Number of people in the vehicle	237	2.194	1.267	1	8
Distance	One-way distance from respondent's Zip code to National Forest	237	120.565	214.908	1	1500
Undeveloped	Visitors who engage in recreational settings such as Wilderness and General Forest Area settings take the value 1 and 0, if they engage in developed recreational settings	237	0.738	0.44	0	1

$$E(y_s, users) = \pi_s [\exp(xs\beta_s)] \quad (9)$$

Consumer surplus per person visit can be calculated as given in (10) (Shaw, 1988),

$$CS/person/visit = -1/\beta_s^{TC} \quad (10)$$

where β_s^{TC} is the coefficient for travel cost variable.

3. Data Description

Data for estimating the Latent Variable (latent class) recreation demand model discussed above were obtained from the U.S. Forest Service's National Visitor Use Monitoring (NVUM) program. The NVUM survey is based on a stratified random sampling design (English et al., 2002). The data were collected from the fourth round, which began in 2012 for a period of 5 years, through 2016. Participating National Forests are sampled every 5 years. During the on-site interviews, information was collected from visitors on their annual number of trips to a sampled National Forest in the last 12 months, and also the number of trips to the sampled National Forest for the activity indicated as the respondent's primary activity.

Information on socio-economic variables was also collected in the NVUM survey, including the gender and age of the respondent. Unlike earlier rounds, primary information on self-reported income and distance was collected for one-third of the sample. Income for the household was recorded as the total annual income of the respondent. Information on travel distance as miles traveled from home to site was also collected from each respondent, and distance from home to substitute location was recorded. Table 1 describes the description of data variables and provides summary statistics of these variables.

Although there is a total of 155 National Forests in the U.S., to facilitate data collection some National Forests were combined within given states resulting in 120 sampling units (forests) for the NVUM survey. In any sampling year during a given 5-year sampling period, on-site survey sampling was conducted for roughly 24 National Forests. For our analysis, a single-site recreational demand model was estimated using data collected for the George

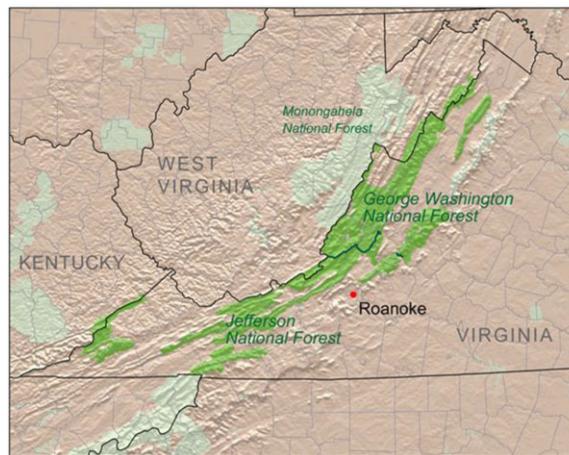


Figure 2. Map of George Washington and Jefferson National Forests.

Washington and Jefferson National Forests. The George Washington and Jefferson National Forests (Figure 2) are located in the southeastern region of the U.S. In 1995, the George Washington National Forest in west central Virginia and the Jefferson National Forest in southwest Virginia were grouped together to form the George Washington and Jefferson National Forests combined management unit. The combined George Washington and Jefferson National Forests management unit (hereafter in this paper referred to as the GW&J National Forest) contains nearly 1.8 million acres of public land with nearly 3 million annual recreation visits. The NVUM survey for GW&J National Forest was conducted in 2015–2016.

4. Empirical Estimation

4.1. Empirical Model for Recreational Demand

The sampling unit for the NVUM survey is a “group,” which can be a single person or a party of persons traveling together, such as a family (Zarnoch, English, and Kocis, 2005). The NVUM survey measures recreation visits to a National Forest on a 12-month basis. Following the TCM protocol, only visitors who were visiting for the primary purpose of recreation were included in our analysis. Our empirical demand equation was specified as,

$$Visits_i = f(TC_i, Income_i, Female_i, Age_i) \quad (11)$$

In (11)², the dependent variable ($Visits_i$) represents the annual number of trips from individual i to the sampled National Forest. Socio-economic variables include annual income ($Income_i$), age (Age_i), and an indicator for a female survey respondent ($Female_i$). $Income_i$ is represented by

²We have not included substitute price variable in our estimation model because of the assumption of weak separability. If recreation demand trips to various sites are separable in the utility function (from other consumption goods), their demands represent a system of demand equations with theoretical cross-equation linkages. That system must be (1) homogeneous of degree zero in travel costs and income (or recreation budget); (2) abide the Cournot and Engel aggregations; and (3) conform to the Slutsky substitution matrix. In regard to the latter, restrictions on substitute price parameters are very strict (the most straightforward interpretation for commonly used semi-log model is that the substitute coefficient must equal zero) (LaFrance, 1990; von Haefen, 2002; Landry et al., 2016). Additionally, we also conducted the likelihood ratio-test to check weak separability assumption in our estimation model. The likelihood ratio test has a null hypothesis that the coefficient on substitute price variable is zero. Our chi-square statistic is 1.827 and we fail to reject the null hypothesis at 5% significance level.

the total annual income of the household. The price of a recreational trip is equal to travel costs for individual i (TC_i) estimated as the sum of driving and time costs following the equation:

$$TC = (2 \times Distance \times \$0.1454/mile)/PeopleVeh + 0.33 \times \frac{Income}{2000} \times \frac{2 \times Distance}{40 mph} \quad (12)$$

In (12), driving costs are a function of one-way distance (*Distance*) from an individual's origin to the destination, the average operating costs (variable costs) per mile for a typical sedan type car in 2016 of 14.54 cents/mile as defined by the American Automobile Association (AAA, 2016), and the number of passengers per vehicle (*PeopleVeh_i*). Time costs are a function of travel time estimated by dividing the round-visit distance by an average speed of 40 mph (Rosenberger and Loomis, 1999) and the opportunity cost of time, which was evaluated at one-third of the wage rate (Baerenklau, 2010). The wage rate was estimated by dividing the income variable per annum by 2000 (Hynes and Greene, 2013). All three variables (round-visit distance, income, and time) are considered exogenous.

4.2. Empirical Model for Membership Probabilities

Earlier estimation of latent count models includes Wedel et al. (1993) and Bockenholt (1993) and more recently, Martinez-Cruz and Sainz-Santamaria (2015) who assume the individual class-membership probability to be a function of only a constant term. We use a more generalized framework where class-membership probabilities are a function of covariates. Following Baerenklau (2010), we modeled the membership probability in (13) empirically as follows:

$$\pi_s = f(Female, Age, Income, Distance, Undeveloped) \quad (13)$$

where *Undeveloped* in (13) is a dummy variable that takes a value one if visitor engages in wilderness or general forest area settings (undeveloped settings) and zero if they engage in developed settings.

5. Results

Model estimation results are summarized in Tables 2, 3, 4, 5, and 6. Table 2 provides the estimation results for the constrained and unconstrained models for one-group and two-group versions based on the on-site Poisson model³. Tables 3 and 4 provide the welfare calculations and the difference in the welfare calculations with bootstrap standard errors, and 95% confidence intervals for estimation results reported in Table 2, respectively. Table 5 provides the estimation results for the three-group version of the on-site Poisson model. Table 6 provides the profile and consumer surplus estimates for the three-group version of the model results reported in Table 5.

³However, when the endogenous variable is overdispersed (i.e., the conditional mean and variance are not equal), then the Poisson model simple parameterization is replaced by a distribution which captures overdispersion. Such models include the Poisson lognormal model (Greene, 2007, p. 8) and the Negative Binomial model (Greene, 2007, p. 5). The difference in these models lies in the distributional assumption of the unobserved factor, ε . We conducted a likelihood ratio test to check if the overdispersion parameter is statistically different from zero. With chi-square value of 5231.17 and Prob > chibar2 = 0.000, we found the presence of overdispersion in our model. We estimated the on-site Poisson model for our analysis. We then estimated an on-site lognormal Poisson model. However, with this new distribution, we did not achieve a global maximum. We increased random starting values and also increased the number of iterations within random starting value sets. This did not improve our results. Therefore, we estimated the on-site Poisson model for our analysis since the Poisson is a member of the linear exponential family and as such provides consistent estimators even under overdispersion.

Table 2. On-site Poisson regression results: one-group and two-group versions

	Model 1	Model 2	Model 3
Class 1 demand			
Female	-0.232 (-0.83)	-0.407 (-1.72)	0.318 (1.38)
Age	0.024** (3.27)	0.019** (2.59)	0.012* (2.08)
Income	-0.002 (-0.48)	-0.002 (-0.64)	-0.006* (-2.17)
Travel cost	-0.080** (-3.14)	-0.040** (-2.72)	-0.075** (-2.84)
Intercept	2.865*** (7.25)	2.507*** (8.39)	4.321*** (11.62)
Class 2 demand			
Female			-0.480 (-1.82)
Age			0.029*** (4.18)
Income			-0.008 (-1.76)
Travel cost			-0.010 (-1.59)
Intercept			1.000* (2.50)
Class 2 membership			
Female			0.714 (1.83)
Age			-0.016 (-1.42)
Income			-0.009 (-1.64)
Distance			0.018* (2.12)
Undeveloped			-0.470 (-1.24)
Intercept			1.278* (2.06)
	N = 237	N = 219	N = 237

Notes: ***coefficient significance at 1%, **coefficient significance at 5%, and *significance at 10%.

We discuss our results in the following ways. First, in Table 2, we compare our constrained models with the unconstrained models in terms of price behavior and consumer surplus estimates. In Table 2, we restrict our comparisons to only the one-class and two-class model versions⁴. We discuss results of the three-class version separately in Table 5, where we discuss the characteristics of different classes of visitors from our three-class models.

Model 1 estimates homogeneous preference without *ad-hoc* truncation (corresponding to the demand curves labeled PC and P'B'C in Figure 1, Panels A and Panel B, respectively) and estimates a one-class on-site Poisson model. Model 2 imposes a constraint on the estimated model, namely *ad-hoc* truncation (corresponding to demand curves labeled PBQ_b in Figure 1, Panel B), and estimates a one-class on-site Poisson model with *ad-hoc* truncation. Model 3 represents

⁴Since many of the socio-economic variables are individually insignificant, we conducted a joint-significance test of socio-economic variables (*Age*, *Gender*, and *Income*) for modeling demand and membership probabilities of visitors to the George Washington and Jefferson National Forest. With chi-square value of 9.38 and 503.74, we find that socio-economic variables are jointly significant at the 5% significance level in modeling demand and membership probabilities, respectively. Additionally, it is common in the literature to include socio-economic variables in demand functions for National Forest recreation trips (Sardana et al. 2016; Nakatani and Sato 2010; Bowker et al. 2009) and membership probability specifications (Baerenklau 2010; Hynes and Greene 2013).

Table 3. Consumer surplus estimates and bootstrap standard errors (replications = 100)⁵

Model	Consumer Surplus	Bootstrap Std. Err.	Z	$P > z$	95% Confidence Interval ⁶
Model 1	12.478	4.730	2.64	0.008	6.625, 25.674
Model 2	24.994	9.639	2.59	0.010	12.224, 49.973
Model 3	34.281	19.527	1.76	0.079	6.185, 82.584

Table 4. Difference in consumer surplus estimates and bootstrap standard errors (replications = 100)

Model	Consumer Surplus	Bootstrap Std. Err.	Bias	95% Confidence Interval
Model 2–1	12.516	6.230	0.745	3.791, 28.166
Model 3–1	21.803	19.530	8.775	6.548, 82.963

the unconstrained model where we relax the assumptions of homogeneous preferences and *ad-hoc* truncation (corresponding to the demand curves labeled P'D and P'Q_c in Figure 1, Panel A) and then estimate a two-class on-site Poisson model without *ad-hoc* truncation⁷.

On-site Poisson model regression results for Models 1–3 are shown in Table 2. In all three models, the travel cost coefficient is statistically significant at the 5% level and negative as expected theoretically. The travel cost coefficient of 0.080 in Model 1 (homogeneous preferences) as compared to 0.075 and 0.010 for Class 1 and Class 2, respectively, in Model 3 (heterogeneous preferences) suggests that, on an average, price responsiveness in Model 1 is higher as compared to price responsiveness for both classes in Model 3, thus resulting in flatter demand curve under Model 1 (e.g., PC in Panel A, Figure 1). The travel cost coefficient of 0.040 in Model 2 as compared to 0.080 in Model 1 suggests that, on an average, price responsiveness in Model 2 (homogeneous with *ad-hoc* truncation) is lower as compared to price responsiveness in Model 1 (homogeneous, no *ad-hoc* truncation), thus resulting in flatter demand curve under Model 1 (e.g., P'B'C in Panel B, Figure 1).

The model estimation results reported in Table 2 were used to calculate the consumer surplus estimates shown in Table 3, which were then used to calculate the difference in consumer surplus estimates shown in Table 4⁸. In comparing Model 1 (homogeneous preferences) with Model 3 (heterogeneous preferences), we quantify the effects of the constraint of homogeneous preferences on consumer surplus estimates. As shown in Table 3, Model 1 resulted in a statistically significant estimate of consumer surplus per individual per trip of about 12 USD. Model 3 resulted in a

⁵We used standardized normal probability plot, to check normality of our bootstrap variables and found them to be non-normal so reported percentile CI instead of normal approximation.

⁶Percentile confidence interval.

⁷Visitors in Class 2 travel more than visitors in Class 1 (given by the positive coefficient of distance in the membership probability).

⁸If the confidence intervals for the means of two independent populations do not overlap, it implies statistically significant difference between the means. However, the opposite is not necessarily true. Confidence intervals can overlap, yet the two means can be significantly different from one another. Therefore, instead of comparing confidence interval for the means of two populations, the confidence interval on the *difference* between the two groups should be estimated (Schenker & Gentleman 2001).

Table 5. On-site Poisson regression results: three-group version of model 3

Variables	Class 1: Local Residents	Class 2: Recreation Enthusiasts	Class 3: Casual Users
Class demand			
Female	0.330*** (7.49)	0.0267 (0.43)	-0.188 (-1.12)
Age	0.009*** (6.85)	0.009*** (5.58)	0.013* (1.97)
Income	0.004*** (4.04)	0.007*** (9.25)	0.005* (2.10)
Travel cost	-0.069*** (-11.92)	-0.016*** (-10.57)	-0.097*** (-7.63)
Intercept	4.552*** (52.86)	2.841*** (27.79)	0.886*** (3.86)
Class membership			
Female		-0.539 (-0.84)	0.427 (0.70)
Age		-0.007 (-0.41)	-0.036* (-2.10)
Income		0.018 (1.52)	0.024* (2.05)
Distance		0.036* (2.24)	0.039* (2.42)
Undeveloped		0.476 (0.74)	-0.421 (-0.69)
Intercept		-0.463 (-0.42)	1.122 (1.06)
N = 237			

Notes: ***coefficient significance at 1%, **coefficient significance at 5%, and *significance at 10%.

Table 6. Trip profile: three-class model

	Local Residents	Recreation Enthusiasts	Casual Users
Sample	0.080	0.313	0.606
Expected trips	113	31	3
Consumer surplus per person per trip	\$14.50	\$62.50	\$10.31

statistically significant estimate of consumer surplus⁹ per individual per trip of about 34 USD¹⁰. Thus, imposing the constraint of homogeneous preferences on the visitors to the GW&J National Forest resulted in about a 64% underestimation of consumer surplus. Also, as shown in Table 4, the difference in the consumer surplus per trip (22 USD) between the two models is statistically significant¹¹.

⁹We bootstrap the difference in consumer surplus for Model 1 and Model 3. Other approaches for calculating the difference include a convolution approach (Poe, Severance-Lossin, and Welsh, 1994).

¹⁰Consumer surplus is given by the expected consumer surplus, given by $(p_1 * CS_1) + (p_2 * CS_2)$, where p_1 and p_2 are probabilities that visitors belong to Class 1 and Class 2, respectively, and CS_1 and CS_2 are consumer surplus for Class 1 and Class 2, respectively.

¹¹Since the 95% CI does not contain the null effect (i.e., zero), which represents the null hypothesis (i.e., no difference in consumer surplus between the groups), we can be 95% confident that the difference in consumer surplus between heterogeneous preferences as compared to the homogeneous preferences, as suggested by the effect estimate (i.e., 21.803), is statistically significant, which provides evidence for rejecting the null hypothesis. In other words, there is difference between the groups (Hespanhol et al., 2019).

Thus, we find evidence that leads us to fail to reject Hypothesis 1 which states that the assumption of homogenous preferences will lead to an underestimation of consumer surplus as compared to the assumption of heterogeneous preference. This intuitively means that with homogeneous preferences, the entire probability weight is assigned to one class of visitors who are highly price responsive. Relaxing the assumption of homogeneity of preferences allocates some of the probability weights to a different, less price responsive, class of visitors. Therefore, consumer surplus is higher in the unconstrained model.

In comparing Model 2 (*ad-hoc* truncation) with Model 1 (no *ad-hoc* truncation), we quantify the effects of *ad-hoc* truncation on consumer surplus estimates. As shown in Table 3, Model 2 resulted in a statistically significant estimate of consumer surplus per individual per trip of about 25 USD compared to about 12 USD per individual per trip for Model 1. Thus, imposing the constraint of *ad-hoc* truncation resulted in an overestimation of consumer surplus by approximately 52%. Also, as shown in Table 4, the difference in the consumer surplus per trip between the two models is statistically significant¹².

The statistically significant difference in consumer surplus estimates across Models 1 and 2 points toward the welfare impacts of *ad-hoc* truncation in our data. Given this statistical difference, we conclude the following about Hypothesis 2 – assuming the constraint of *ad-hoc* truncation results in the overestimation of consumer surplus in the constrained model. This implies that the visitor group, whose demand is cut-off due to *ad-hoc truncation*, has a relatively smaller consumer surplus in the unconstrained model (area II + III + IV). For example, including these visitors who take more frequent visits and have relatively smaller consumer surplus per trip makes the unconstrained demand curve flatter (P'B'C in Panel B), price responsiveness higher, and consumer surplus lower. Thus, if these visitors are truncated out of the sample, consumer surplus will be overestimated.

Our final model, Model 4, represents the three-class model for the three class version of the one-site Poisson model with three distinct classes of visitors: 1) a group of visitors termed “residents” (Class 1); 2) a distinctly different group termed “recreational enthusiasts” (Class 2); and 3) a group of visitors termed “casual users” (Class 3). The classes are defined based on trip frequency, price responsiveness, and consumer surplus per person. Estimation results for Model 4 are shown in Table 5. For all three classes, the negative and significant estimated coefficients for own travel costs indicate that the number of trips is inversely related to own travel costs for each class, implying a downward-sloping demand curve. An increase in travel costs by 1 USD reduces trips by 6.9%, 1.6%, and 9.7% for local residents, recreational enthusiasts, and casual users, respectively.

The price responsiveness as given by the travel cost coefficient estimates of respective groups in Table 5 is highest for casual users followed by local residents and lowest by recreational enthusiasts. According to economic theory, if a recreational trip to a National Forest is a normal good, we would expect an increase in income to have a positive marginal effect on trips demanded. The positive and significant estimated coefficients for income in Model 4 imply that an increase in income by 1 USD increases trips by 0.4%, 0.7%, and 0.5% for local residents, recreational enthusiasts, and casual users, respectively. The coefficient for age is positive and statistically significant for all the three classes. The positive and statistically significant coefficient for the gender variable (*Female*) for local residents suggests that for this population female respondents, on average, take 33% more trips *ceteris paribus* than male respondents.

Though, in the NVUM data, we do not have data on substitute visits to state/local parks, price responsiveness of casual users intuitively implies that casual users substitute their visits to

¹²Since the 95% CI does not contain the null effect (i.e., zero), which represents the null hypothesis (i.e., no difference in consumer surplus between the groups), we can be 95% confident that the difference in consumer surplus between *ad-hoc* truncation as compared to no *ad-hoc* truncation, as suggested by the effect estimate (i.e., 12.516), is statistically significant, which provides evidence for rejecting the null hypothesis. In other words, there is difference between the groups.

National Forest recreation sites with other recreational facilities (may be local or state parks) as compared to recreational enthusiasts. Based on their engagement in day trips that may be more in the nature of a convenient site for casual day-use visits, the casual user may decide at the spur of the moment to go, say, to a local town park instead. Our definitions of these two types (classes) of visitors and findings in terms of price responsiveness and consumer surplus are consistent with Baerenklau (2010).

Table 6 shows that local residents tend to visit more frequently as compared to recreational enthusiasts and casual users. Local residents, on average, take 113 annual visits as compared to recreational enthusiasts who on an average take 31 annual visits and casual users who take three annual visits. Recreational enthusiasts value each visit more than local residents and casual users. Their consumer surplus per person per trip is about 62 USD as compared to 14 USD for local residents and 10 USD for casual users. Casual users, recreation enthusiasts, and local residents constitute about 61%, 31%, and 8% of total visitors, respectively.

We further use group membership coefficients estimated in Table 5 to define the characteristics of different classes of visitors¹³. The statistically significant positive one-way travel distance coefficient in the membership probability for the recreational enthusiasts and casual users suggests that these visitors travel greater distances to visit the GW&J National Forest as compared to local residents, which is likely attributable to the endogenous spatial sorting discussed by Baerenklau (2010)¹⁴. If an individual travels an additional mile of one-way distance, it increases the relative probability of belonging to the class of recreational enthusiasts as compared to local resident by 0.036.

Descriptive statistics corroborate these membership model estimation results, showing, for example, that the average one-way distance traveled by recreational enthusiasts is about 55 miles compared to about an average of 19 miles by local residents. The negative coefficient on the settings variable suggests that local residents engage more in undeveloped activities as compared to casual users. However, this coefficient estimate was not statistically significant.

The estimation results in Table 5¹⁵ suggest that local residents and casual users differ in terms of their price-responsive behavior and trip frequency. Casual users travel less frequently than local residents. Casual users on an average take three annual visits. They live further away from the GW&J National Forest as compared to local residents and visit developed settings such as day-used settings as compared to local residents who visit undeveloped settings. Casual users have a flatter demand curve than local residents and thus their price responsiveness is higher (40% more) than local residents. As compared to local residents, casual users derive lower consumer surplus per visit.

6. Implications and Conclusions

In this paper, we compare the effects of *ad-hoc* truncation and homogeneous preference constraints on recreation demand and welfare measure estimates associated with trip to the GW&J National Forest in the southeastern U.S. In most cases, observations that are dropped due to constraints imposed on the definition of visitors constitute a relatively small proportion

¹³Baerenklau (2010) suggests that membership probability is a function of all socio-economic variables. Thus, we conducted a joint-significance test of all socio-economic variables for modeling membership probability. With chi-square value of 12.87 and P value = 0.025, we find that socio-economic variables are jointly statistically significant at 5% level in explaining membership probabilities.

¹⁴Also, as pointed out by an anonymous reviewer, the place of residence is endogenous. People may want to live in this area, precisely because of the proximity to the forest and its outdoor recreation opportunities.

¹⁵Information criterion based on CAIC shows that the model with three classes performs better than the model with two classes. CAIC imposes an additional sample size penalty on the log-likelihood as compared to the frequently used AIC criterion, therefore favoring more parsimonious models (Wedel et al. 1993). CAIC statistic is 9,847.25, 5,484.6469, 3,855.1835, and 2,195.9822 for Model 1, Model 2, Model 3, and Model 4, respectively.

of the visitor population but have a significant impact on consumer surplus estimates. In our analysis, we found that differences in consumer surplus per person per trip are contingent on assumptions regarding visitor preferences (homogenous vs. heterogeneous), the elasticity (slope) of the demand curve constrained by homogenous preferences, and the binding *ad-hoc* truncation point set by the researcher.

In general, we found that the assumption of homogenous preferences resulted in the underestimation of consumer surplus, and right-hand side *ad-hoc* truncation resulted in the overestimation of consumer surplus. Thus, these constraints could lead to inaccurate (biased) estimates of the benefits of recreation sites such as the GW&J National Forest. Inaccurate (biased) estimates of benefits, in turn, could lead to incorrect policy and management decisions based on, for example, benefit-cost analysis.

We also found that consumers surplus estimates are sensitive to different types or classes of visitors based on distance traveled to the recreation site (GW&J National Forest) and the particular forest setting visited (e.g., developed site vs. undeveloped site). These results imply that visitors can be segmented into different user-groups or “markets” based on different visitor (and visit) characteristics and preferences.

The ability to segment visitors into different types or classes can facilitate differential pricing policies, should that be something natural resource management agencies may want to consider in cases where user fees (e.g., entrance fee, parking fee) are charged for access to recreation sites. For example, our results show that recreation enthusiasts have a substantially higher consumer surplus (willingness-to-pay) per person per trip as compared to casual users and local residents. Also, casual users have a slightly higher consumer surplus per person per trip as compared to local residents. Thus, from both economic efficiency and equity perspectives, natural resource management agencies may want to consider charging a different user fee per trip or annual user fee to different types of users such as local residents, casual users, and recreational enthusiasts identified by the latent variable (latent class) analysis in this study. For example, local residents, who are also the most frequent users, might be charged a relatively low “local resident” daily or annual visit fee as compared to recreation enthusiasts who are non-local residents and travel less frequently to the site from greater distances.

Acknowledgements. We would like to gratefully acknowledge Professor J. Scott Shonkwiler and Professor Kevin J. Boyle for their detailed comments on an earlier draft of the manuscript.

Financial Support. This work was supported by the U.S. Department of Agriculture (U.S.D.A) through the Georgia Agricultural Experiment Station under the W-4133 Regional Research Project entitled “Costs and Benefits of Natural Resources on Public and Private Lands: Management, Economic Valuation, and Integrated Decision-Making” and by the Southern Research Station, U.S.D.A Forest Service.

References

- American Automobile Association (AAA).** Your Driving Costs, 2016. Internet site: <https://exchange.aaa.com/wp-content/uploads/2017/05/2016-YDC-Brochure.pdf> (Accessed March 18, 2020).
- Baerenklau, K.A.** “A Latent Class Approach to Modeling Endogenous Spatial Sorting in Zonal Recreation Demand Models.” *Land Economics* **86**,4(2010):800–16.
- Böckenholt, U.** “A Latent-Class Regression Approach for the Analysis of Recurrent Choice Data.” *British Journal of Mathematical and Statistical Psychology* **46**,1(1993):95–118.
- Borzykowski, N., A. Baranzini, and D. Maradan.** “A Travel Cost Assessment of the Demand for Recreation in Swiss Forests.” *Review of Agricultural, Food and Environmental Studies* **98**,3(2017):149–71.
- Bowker, J.M., C.M. Starbuck, D.B.K. English, J.C. Bergstrom, R.S Rosenburger, and D.C. McCollum.** “Estimating the Net Economic Value of National Forest Recreation: An Application of the National Visitor Use Monitoring Database.” Faculty Series Working Paper, FS 09-02, September 2009. Athens, GA: The University of Georgia, Department of Agricultural and Applied Economics, 2009. Internet site: <http://ageconsearch.umn.edu/handle/59603> (Accessed 12 December 2020).
- Bozdogan, H.** “Model Selection and Akaike’s Information Criterion (AIC): The General Theory and Its Analytical Extensions.” *Psychometrika* **52**,3(1987):345–70.

- Dempster, A.P., N.M. Laird, and D.B. Rubin.** "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39,1(1977):1–22.
- Egan, K., and J. Herriges.** "Multivariate Count Data Regression Models with Individual Panel Data from An On-Site Sample." *Journal of Environmental Economics and Management* 52,2(2006):567–81.
- Englin, J., and J.S. Shonkwiler.** "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation." *The Review of Economics and Statistics* 77,1(1995):104–12.
- English, D., et al.** *Forest Service National Visitor Use Monitoring Process*. Asheville, NC: USDA Forest Service Southern Research Station, General Technical Report SRS-57, 2002.
- Freeman, A.M.** *The Measurement of Environmental and Resource Values: Theory and Methods*. No. GTZ-1574. Resources for the Future, 1992.
- Greene, W.** "Functional Form and Heterogeneity in Models for Count Data." *Foundations and Trends® in Econometrics* 1,2(2007):113–218.
- Haab, T.C., and K.E. McConnell.** *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Northampton: Edward Elgar Publishing, 2002.
- Heckman, J., and B. Singer.** "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica: Journal of the Econometric Society* 52(1984):271–320.
- Hellerstein, D.M.** "Using Count Data Models in Travel Cost Analysis with Aggregate Data." *American Journal of Agricultural Economics* 73,3(1991):860–66.
- Hespanhol, L., C.S. Vallio, L.M. Costa, and B.T. Saragiotto.** "Understanding and Interpreting Confidence and Credible Intervals Around Effect Estimates." *Brazilian Journal of Physical Therapy* 23,4(2019):290–301.
- Hynes, S., and W. Greene.** "A Panel Travel Cost Model Accounting for Endogenous Stratification and Truncation: A Latent Class Approach." *Land Economics* 89,1(2013):177–92.
- LaFrance, J.T.** "Incomplete Demand Systems and Semilogarithmic Demand Models." *Australian Journal of Agricultural Economics* 34,2(1990):118–31.
- Landry, C.E., A.R. Lewis, H. Liu, and H. Vogelsong.** "Economic Value and Economic Impact of Visitation to Cape Hatteras National Seashore: Addressing Onsite Sampling." *Marine Resource Economics* 31,3(2016):301–22.
- Martinez-Cruz, A.L., and J. Sainz-Santamaria.** "Recreational Value of Two Peri-Urban Forests in Mexico City." *El Trimestre Economico* 117(2015):1–26.
- Nakatani, T., and K. Sato.** "Truncation and Endogenous Stratification in Various Count Data Models for Recreation Demand Analysis." *Journal of Development and Agricultural Economics* 2,8(2010):293–302.
- Parsons, G.R.** "A Note on Choice of Residential Location in Travel Cost Demand Models." *Land Economics* 67,3(1991):360–64.
- Poe, G.L., E.K. Severance-Lossin, and M.P. Welsh.** "Measuring the Difference (X—Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76,4(1994):904–15.
- Rosenberger, R.S. and J.B. Loomis.** "The Value of Ranch Open Space to Tourists: Combining Observed and Contingent Behavior Data." *Growth and Change* 30,3(1999):366–83.
- Sardana, K., J.C. Bergstrom, and J.M. Bowker.** "Valuing Setting-Based Recreation for Selected Visitors to National Forests in the Southern United States." *Journal of Environmental Management* 183(2016):972–9.
- Schenker, N., and J.F. Gentleman.** "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals." *The American Statistician* 55,3(2001):182–6.
- Shaw, D.** "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37,2(1988):211–23.
- U.S. Forest Service National Visitor Use Monitoring Survey Results National Summary Report**, 2016. Internet site: <https://www.fs.fed.us/recreation/programs/nvum/pdf/5082016NationalSummaryReport062217.pdf> (Accessed March 18, 2020).
- Von Haefen, R.H.** "A Complete Characterization of the Linear, Log-Linear, and Semi-Log Incomplete Demand System Models." *Journal of Agricultural and Resource Economics* 27,2(2002):281–319.
- Wedel, M., W.S. DeSarbo, J.R. Bult, and V. Ramaswamy.** "A Latent Class Poisson Regression Model for Heterogeneous Count Data." *Journal of Applied Econometrics* 8,4(1993):397–411.
- Zarnoch, S.J., D.B.K. English, and S.M. Kocis.** "An Outdoor Recreation Use Model with Applications to Evaluating Survey Estimators." Res. Pap. SRS-37, Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station 15, p. 37, 2005.

Cite this article: Sardana K, Bergstrom JC, and Bowker JM (2021). Effects of *Ad-hoc* Data Truncation and Homogeneous Preferences on Recreational Demand and Values: An Application to the George Washington and Jefferson National Forests. *Journal of Agricultural and Applied Economics* 53, 153–167. <https://doi.org/10.1017/aae.2020.30>