Modeling Demand for Outdoor Recreation Settings with Choice Based Data Accounting for Exogenous and Endogenous Stratification

DRAFT - COMMENTS WELCOMED

Kavita Sardana (presentor)

Graduate Research Assistant

Department of Agricultural and Applied Economics

The University of Georgia, Athens

J. Michael Bowker

Research Social Scientist

Southern Research Station

USDA Forest Service, Athens GA

John C. Bergstrom

Russell Professor of Public Policy and Professor of

Agricultural and Applied Economics

The University of Georgia, Athens

Selected Paper prepared for presentation at the Agricultural and Applied Economics Association 2010 AAEA,CAES,and WAEA Joint Annual Meeting, Denver, Colorado, July 25-27, 2010

Abstract

Estimating regional demand models by pooling different samples without correcting for such differences causes model misspecification as each sample belongs to a different population. Weighted regression using Pseudo likelihood to account for differences in sample population with adjustment for heteroskedasticity improves efficiency but the estimates are biased. We estimate regional demand for National Forest settings types in the southeastern states of U.S using weighted and unweighted regression. Using estimation of demand for National Forests as a case study, we resolve problems relating to inference about the data generating process when different samples are pooled together. We show that though efficiency of weighted estimates improves after correcting for heteroskedasticity, they still remain biased as the weights interact with covariates to explain part of model misspecification. In this paper, we show that it is best to use unweighted regression including interactions with weights as covariates.

Introduction

Many if not all on-site samples are choice based samples. In a choice-based sample, stratification is on the endogenous variable, directly affecting the kernel of the likelihood. Econometric procedures used in estimation therefore needs to account for endogenous stratification in order to obtain consistent parameter estimates (Manski and McFadden, 1981). This is achieved by deriving appropriate weights for the relevant distribution in a weighted regression. For a count outcome, Shaw (1998) and Englin and Shonkwiler (1995) derive weights to correct for endogenous stratification for poison and negative binomial distributions respectively. However, when stratification is on the exogenous variable, estimation proceeds in a regular fashion. In this case, the econometric correction amounts to adding a constant of proportionality which does not affect the kernel of the likelihood. Manski and McFadden(1981) point out that it is important for practitioners to understand that in exogenous stratification, distribution of strata is defined on the domain of exogenous variables. In that case knowledge about the distribution of exogenous covariates alone is sufficient to know the distribution of strata, even if the distribution of strata affects the choice probability only trivially. Wooldridge (2001) shows that under the assumption of homoskedasticity, econometric procedures do not need to account for exogenous stratification. He further shows that weighted estimates that correct for exogenous stratification are consistent but less efficient than un-weighted estimates in a linear specification.

For the purpose of inference about the relevant population, weighted regression is often used in empirical estimation to correct for differences in sampling rates due to exogenous variables, such as race, age, gender etc. For example, suppose there are 50 % females in the relevant population and the sample includes only 30 % females. In this case, weighting is used to equate the sampling distribution to the population distribution by using weights usually derived from the US Census. Even in that case differences in weighted and unweighted results point to some form of model misspecification. Korn and Graubard (1995) give at least two reasons for the weighted estimates to be different from unweighted estimates when stratification is on the exogenous variable; the model must be much misspecified or an omitted variable must have a strong interaction with the independent variable and must be highly correlated with the weights. Winship and Radbil (1994) attributes the differences in the weighted and unweighted results to pooling two different samples together. This is particularly relevant in regional models when observations are pooled together due to insufficient data. Another possible reason for the differences is that the weights when interacted with covariates account for the omitted variables in the regression. DuMouchel and Duncan(1983) gives a simple F test to test the later reason.

The objective of this paper is to shed some insight on the reasons for differences in parameter estimates of weighted and unweighted regression especially in estimating models where different samples are pooled together. We show how consistent and efficient parameter estimates can be obtained if that is the case. This information is of relevance to the federal agencies such as the USDA Forest Service. The USDA Forest Service conducts on-site samples of recreation visitor use on a regular basis for the purposes of projections and budget allocation. For the purposes of low survey costs, they are more interested in regional estimates than individual forest estimates. In the paper we empirically estimate demand for National Forest in the southeastern states of the U.S. for settings types. We also show how inference can be completely erroneous if incorrect specification for standard errors is used.

This paper is organized as follows. In the first section we briefly discuss the theory of weighted and unweighted regression. In the second section we specify the empirical models of demand for 4 settings type- Day-Used Developed Sites (DUDS), Overnight-Used Developed Sites (OUDS) ,General Forest Area (GFA) and Wilderness(WILD). In the third section we explain our results followed by conclusions. Appendix contains 4 tables containing summary statistics for each settings types and the last table in appendix contains weighted means for all settings types.

Data Description

Data for estimating the empirical model specified above were obtained from the National Visitor Use Monitoring Program (NVUM). NVUM started collecting visitor use information for a stratified on-site sample in the year 2000. In its first four year cycle (2000-2003), NVUM collected information on annual number of visits to National Forest for the primary purpose of outdoor recreation, primary activity for an individual, and other socio-economic variables. Information on home zip code for individuals was collected for the calculation of implicit price variable (Travel Cost) and to use IRS data (available according to zip code) as a proxy for the income variable. The original master dataset has information on 10 RPA regions and 120 National Forests across the U.S. For further information on adjustments made in the original dataset refer to Bowker et al (2009). NVUM is based on a stratified sample technique suggested by English (2002). Every National Forest within the sample is divided into 12 strata according to site-type and site use. Site types or settings include Day Used Developed Sites (DUDS), Overnight Used Developed Sites (OUDS), General Forest Area (GFA) and Wilderness (WILD). Site use includes Low (L), Medium (M) and High (H) usage. Random samples are drawn from each stratum. For the analysis in this case, we use data for the southeatern U.S. or U.S. Forest Service Region 8. The data is collected for 14 National Forests including the Chattahoochee-Oconee National Forest, George Washington-Jefferson National Forest, Croatan National Forest, Daniel Boone National Forest, Cherokee National Forest, Francis Marion National Forest, Conecuh National Forest, Ozark National Forest, Apalachicola National Forest, DeSoto National Forest, Ouachita National Forest, Bienville National Forest, Kisatchie National Forest, Davy Crockett National Forest, and Land between Lakes National Forest. The NVUM survey sampled 25% of total National Forests in its 2000 cycle and 20% in its Oct 2004 cycle. The dataset for southeastern region include 7000 sample observations.

Theoretical Model

When different samples are pooled together, estimation can proceed using Pseudo likelihood, first used by Besag (1975; 1977).A Pseudo likelihood estimation is based on the assumption that each random process is independent. In the case of regional demand models, demand for various samples across the region are independent of each other.
We briefly explain the methodology below from Wang et al. (2004).
let,

$$X = (X_1, X_2, ..., X_n) \tag{1}$$

be random variables with probability density functions

$$f_1, f_2, ..., f_n \tag{2}$$

The density of interest is

$$f(., \theta), \theta \in \Theta \tag{3}$$

of a study variable X. At least in some qualitative sense, the

$$f_1, f_2, ..., f_n$$

is thought to be like

$$f(., \theta)$$

We assume that each independent distribution is related to the distribution of interest through relevant weights. Pseudo likelihood or what is popularly known as Power likelihood is therefore given by,

$$\prod_{j=1}^{m}\prod_{i=1}^{nj} f^{\lambda_j}(x_{ij}, \theta) \tag{4}$$

where, $j = 1, 2, ...m$ are the number of independent random samples, and $i = 1, 2, ...nj$ are the number of individuals in each sample. Therefore, the concept of pseudo likelihood is used to estimate the parameter of interest. It is important to understand that the weights, though constructed based on exogenous variables, do not enter the likelihood as a constant of proportionality. Therefore, weights in this case affect the kernel of the likelihood.

In our model, we assume that the data generating process follows a negative binomial distribution correcting for endogenous stratification and truncation.The log likelihood for a negative binomial distribution accounting for truncation and endogenous stratification is given by,

$$log(y)+log\Gamma(y+\alpha^{-1})+ylog(\alpha)+(y-1)(x\beta)-(y+\alpha^{-1})log(1+\alpha exp(x\beta))-log\Gamma(\alpha^{-1}) \tag{5}$$

The score function is given by,

$$\frac{dlogL}{d\beta} = \sum_i (y_i - 1)X_i - (y_i + \alpha^{-1})\frac{\alpha X_i EXP(X_i\beta)}{1 + \alpha X_i EXP(X_i\beta)} \tag{6}$$

The Information matrix is given by the inverse of the second derivative,

$$\frac{dlogL}{d\beta'\beta} = \sum_i \frac{-(y_i + \alpha^{-1})\alpha X_i' X_i EXP(X_i\beta)}{(1 + \alpha X_i EXP(X_i\beta))'(1 + \alpha X_i EXP(X_i\beta))} \tag{7}$$

The score function for weighted regression is given by,

$$\frac{dlogL}{d\beta} = \sum_j \lambda_j \sum_i (y_{ij} - 1)X_{ij} - (y_{ij} + \alpha^{-1})\frac{\alpha X_{ij} EXP(X_{ij}\beta)}{1 + \alpha X_{ij} EXP(X_{ij}\beta)} \tag{8}$$

If we make an assumption of a power likelihood.

The Information matrix is given by the inverse of the second derivative,

$$\frac{dlogL}{d\beta'\beta} = \sum_j \lambda_j \sum_i \frac{-(y_{ij} + \alpha^{-1})\alpha X_{ij}' X_{ij} EXP(X_{ij}\beta)}{(1 + \alpha X_{ij} EXP(X_{ij}\beta))'(1 + \alpha X_{ij} EXP(X_{ij}\beta))} \tag{9}$$

Empirical Model

In a stratified sample, sampling weights are used to expand each individual to be representative of the proper population. It is given by,

$$\frac{N_j}{n_j}$$

where, N_j are the number of individuals in stratum j in the population and n_j are the number of individuals sampled in stratum j. In many cases,the numerator is known. But in cases where it is not known, it needs to be estimated. In the case of NVUM,N_j is not observed directly and is estimated by,

$$NVEXPAND_{ji} = ExitingTraffic_{ji} * PropLastExit_{ji}$$

where exiting traffic is the average exiting traffic count per day for the stratum and proportion last exit is the ratio of last exiting recreation vehicles to total count of vehicles.

The NVUM survey sample collects sufficient data to allow computation of weights. Its computation is based on the proportion last exited visitors in a given stratum in a forest. These weights are used in weighted regression. For further information on NVUM survey samples refer to Appendix B in Bowker et al(2009).

EmpiricalModel Specification

We model visits to a National Forest as a truncated negative binomial model correcting for endogenous stratification. We estimate both weighted and unweighted regional demand models for settings using the following empirical specification:

$$\begin{aligned} NFV12MO = f(&PEOPVEH, GENDER, AGE, TC, HF, OSITES, \\ &OVERNTE, ECOREG, SUPPLYVAR) \end{aligned}$$

The dependent variable is the number of annual recreation visits to a National Forest per individual/group. Demand for visits is a function of: own price (TC), number of people in the vehicle (PEOPVEH), annual income (INCOME), gender (GENDER1), age (AGE), and an indicator for staying overnight (ONITE), an indicator if an individual visited any other site (OSITES). a dummy variable if forest belongs to subtropical ecoregion (SUBTROP), a dummy variable if a forest belongs to hot continental ecoregion (HOTCONT) and a dummy variable if a forest belongs to mountain ecoregion (MOUNTAIN). In the model we drop the dummy for subtropical ecoregion. An additional term has been incorporated to capture the differences between high and low frequency users (HF), where HF=1 if number of annual visits was greater than 15, else zero. The supply variables for the General Forest Area setting include percentage of forest area with-in a radius of 100miles of origin (FORESTP) and miles of trails in a National Forest as a proxy for access to general forest areas (TRAILS). Supply variables for the Overnight Used Developed Sites settings include total number of tent camping sites in a National Forest (TENTC) and total number of establishments in recreation and vacation camps category with-in a 100 miles of origin(SUMCAMPS). Supply variables in Day Used Developed Sites include total number of recreation areas in a National Forest with picnic tables as a proxy for total number of day use sites (PICNICTAB), total number of recreation areas in a National Forest with swimming areas as a proxy for high-attraction day use sites (SWIMMING) and total number of establishments in nature parks and similar institutions with-in a 50miles of origin as a proxy for private day used sites (SUMNATPARK). Supply variables for the Wilderness setting include miles of designated wilderness area in a given National Forest(DESIG).

Results

Tables 1 through 4 include results for the settings types of GFA, DUDS, OUDS and WILD. The first row gives the coefficient for weighted and un-weighted models referred to as Model1 and Model 2, respectively. The second row gives the standard errors computed using the Newton-Raphson algorithm assuming homoskedasticity, and the third row includes White's standard errors corrected for heteroskedasticity. The purpose of including heteroskedasticity corrected standard errors is to show that though in the un-weighted regression, assumption of homoskedasticity can be maintained, in the case of weighted regression, the same cannot be assumed. This result confirms the claim made by Winship and Radbill(1994). This is because covariates in the weighted regression become correlated with the error term. It is therefore important to correct weighted standard errors for heteroskedasticity.We will explain this later when we discuss our results in Table 5 .

Table 1: General Forest Area

	Model1	Model2
Intercept	1.282	0.968
	(0.130)*	(0.001)*
	(.160)*	(.258)*
HOTCON	0.197	-0.077
	(.0006)*	(.006)*
	(.058)*	(.093)
MOUNTN	0.113	0.131
	(.0006)***	(.0006)*
	(.057)**	(.082)
FOREST	0.003	0.007
	(.002)***	(0.00002)*
	(.002)	(.004)***
TRAIL	-0.0004	0.0001
	(.0001)*	(0.000001)*
	(.0001)*	(.0001)
INCE	-0.000009	-0.000004
	(0.00002)*	(0.00000003)*
	(.000003)*	(.000004)
AGE	0.002	0.003
	(0.001)	(0.00001)*
	(.001)	(.0021)***
GENDER	-0.164	-0.113
	(0.046)*	(.0005)*
	(.053)*	(.101)
PEOPVEH	-0.027	-0.042
	(0.013)**	(.0001)*
	(.013)**	(.021)**
OSITE	-0.067	-0.026
	(0.040)***	(0.0005)*
	(.042)***	(.071)
OVERNTE	0.039	0.193
	(0.041)	(0.0005)*
	(.042)	(.063)*
TC	-0.003	-0.003
	(0.003)*	(0.000003)*
	(.0004)*	(.0007)*
HF	1.816	1.707
	(0.034)*	(0.0003)*
	(.031)*	(.050)*
ALPHA	0.561	0.425
	(0.047)*	(0.0003)*
	(.046)*	(.054)*
NOBS	1979	
LOGL	54920.7	480453000
BIC	-54867.5	-480453000
*1% significance		
**5% significance		
***10% significance		

Results in Table 1 show that in explaining demand for trips to General Forest Area setting, in Model1 standard errors with heteroskedasticity correction are bigger but do not change inference in terms of significance of the coefficient. Such is not the case with unweighted regression. The reason for a change in significance of coefficients is two fold. Not only are the heteroskedasticity corrected standard errors significantly different but the coefficient estimates become inconsistent due to significant interactions of some important variables with the weights. This can be seen from Table 5. These variables include a dummy for hot continental and mountain ecoregion, income variable, a dummy for overnight stay, and supply variable, trails. Difference in signs of weighted and unweighted models can be attributed to inconsistency of the weighted model.

Results in Table 2 show that for the Day Used Developed sites regression, the intercept and the dispersion parameter both become insignificant in the weighted regression. An insignificant dispersion parameter points to the failure of an important theoretical assumption of the model; i.e. the difference in mean and variance of the population.This points to inconsistency of parameter estimates of weighted regression.

Results in Table 3 show that in explaining demand for trips to Overnight Use Developed sites, a dummy for overnight stay changes sign from positive

Table 2: Day Used Developed Sites

	Model1	Model2
Intercept	0.673	-0.585
	(0.176)*	(0.007)*
	(.237)*	(.846)
HOTCONT	0.095	0.353
	(.073)	(0.002)*
	(.069)	(.125)*
MOUNTAIN	-0.257	0.178
	(.105)**	(.003)*
	(.103)**	(.201)
PICNICTAB	0.001	-.0004
	(.00008)	(0.00002)*
	(.00008)	(.0002)**
NATPARK	.004	0.0001
	(.003)	(0.000005)*
	(.004)	(.007)
SWIMMING	-0.019	-.066
	(.014)	(.0003)*
	(.014)	(.025)*
INCE	-0.0001	-0.000009
	(0.00002)*	(0.00000007)*
	(.000004)*	(.0000056)***
AGE	-0.003	0.005
	(0.001)**	(0.00003)*
	(.002)***	(.003)***
GENDER	-0.012	-0.039
	(0.047)	(.001)*
	(.050)	(.096)
PEOPVEH	-0.053	-0.039
	(0.014)*	(.0003)*
	(.0144)*	(.0308)
OSITE	-0.259	-0.017
	(0.0409)*	(0.001)*
	(.055)*	(.111)
OVERNTE	-0.193	-0.439
	(0.073)*	(0.002)*
	(.078)*	(.154)*
TC	-0.003	-0.003
	(0.003)*	(0.000005)*
	(.001)*	(.001)*
HF	2.217	2.248
	(0.050)*	(0.001)*
	(.037213)*	(.072)*
ALPHA	2.592	6.466
	(0.476)*	(0.047)*
	(.652)*	(5.557)
NOBS	2394	
LOGL	36023.8	87280200
*1% significance		
**5% significance		
***10% significance		

Table 3: Overnight Used Developed Sites

	Model1	Model2
Intercept	0.630	1.503
	(0.165)*	(0.003)*
	(.194)*	(.330)*
HOTCONT	-0.370	0.271
	(.118)*	(0.002)*
	(.116)*	(.209)
MOUNTAIN	-0.234	-0.162
	(.081)*	(.002)*
	(.083)*	(.200)
TENTC	0.0003	0.0001
	(.0001)**	(0.00002)*
	(.0001)**	(.0002)
SUMCAMPS	.001	0.0005
	(.001)	(0.00002)*
	(.001)	(.002)
INCE	-0.0001	-0.00003
	(0.000003)*	(0.00000008)*
	(.000003)*	(.00001)*
AGE	0.005	0.004
	(0.002)*	(0.00004)*
	(.002)*	(.003)
GENDER	-0.106	-0.115
	(0.052)**	(.001)*
	(.0589)**	(.119)
PEOPVEH	-0.043	-0.010
	(0.017)*	(.0004)*
	(.0167)*	(.036)
OSITE	-0.196	-0.245
	(0.050)*	(0.001)*
	(.0520)*	(.101)*
OVERNTE	0.009	-0.305
	(0.049)*	(0.001)*
	(.050)	(.097)*
TC	-0.003	-0.002
	(0.0004)*	(0.00001)*
	(.001)*	(.001)*
HF	2.188	1.885
	(0.059)*	(0.001)*
	(.0412)*	(.127)*
ALPHA	1.461	.407
	(0.215)*	(0.001)*
	(.240)*	(.096)*
NOBS	1707	
LOGL	18949.7	31155200
BIC	-18897.6	-31155200
*1% significance		
**5% significance		
***10% significance		

to negative in the weighted regression. Theory suggests a positive sign for the dummy variable for overnight stay in explaining demand for trips to Overnight Use Developed sites. If a visitor stays overnight in a National Forest, demand for overnight use developed site increases. A negative sign for the overnight stay dummy in the weighted regression points to inconsistency of weighted regression.

Table 4 gives the coefficient and standard errors for the wilderness model. In Table 5 only the intercept and income have significant interactions with the weight variable and the interactions with the other covariates of the model are insignificant. In these results, unlike the previous models the signs for weighted and unweighted models stay the same.

In Table 5, we have only included the covariates interacted with weights as the remaining coeffecient remain the same in the unweighted regression.Table 5 shows that weights that are constructed to provide correction for differences in sample rates have strong interactions with covariates included in the model. Therefore, weights interact with covariates to partially or fully explain the variables omitted from the model. This causes errors to be heteroskedatic and if not corrected would result in wrong inference.

Table 4: Wilderness

	Model1	Model2
Intercept	0.481	2.113
	(0.446)	(0.009)*
	(.526)	(.430)*
HOTCONT	0.521	0.374
	(.102)*	(0.004)*
	(.103)*	(.200)***
MOUNTAIN	0.110	-0.258
	(.200)	(.006)
	(.190)	(.335)
DESIGW	0.000005	0.000007
	(.000003)***	(0.0000001)*
	(.000003)	(.000005)
SUMWILDERN	.007	0.007
	(.004)	(0.00007)
	(.002)*	(.002)*
INCE	-0.0002	-0.00003
	(0.000004)*	(0.0000002)*
	(.000007)***	(.00001)*
AGE	-0.004	-0.002
	(0.004)*	(0.0001)*
	(.004)	(.006)
GENDER	-0.077	-0.362
	(0.101)*	(.003)*
	(.100)	(.172)**
PEOPVEH	-0.068	-0.075
	(0.038)***	(.001)*
	(.035)**	(.060)
OSITE	-0.325	-0.161
	(0.105)***	(0.004)*
	(.099)*	(.166)
OVERNTE	-0.192	-0.545
	(0.112)***	(0.004)*
	(.114)***	(.173)*
TC	-0.003	-0.004
	(0.0004)*	(0.00002)*
	(.001)*	(.001)*
HF	2.092	2.337
	(0.145)*	(0.004)*
	(.001)*	(.225)*
ALPHA	3.711	0.873
	(1.735)**	(0.006)*
	(1.869)**	(.381)**
NOBS	618	
LOGL	4059.37	7875890
BIC	-4014.39	-7875890
*1% significance		
**5% significance		
***10% significance		

Table 5: Coefficients of Interactions with Weights

	GFA	DUDS	OUDS	WILD
INTERCEPT	-.146E-04*	-.952E-04**	.315E-04	.572490E-03*
HOTCONT	-.901E-05*	.281E-04	.743E-04*	.228631E-04
MONUNTAIN	.779E-05***	.766E-04**	.794E-05	-.257495E-04
INCE	.715E-09 *	.414E-09	-.217E-08*	-.156295E-07*
AGE	.104E-06	.143E-05*	.370E-06	.224673E-05
GENDER	.200E-06	-.821E-05	-.213E-04***	.224673E-05
PEOPVEH	-.161E-05	.275E-05	.482E-06	-.341661E-04
OSITE	.670E-05	.218E-04***	-.127E-04	.220476E-04
OVERNTE	.127E-04**	-.687E-04	-.353E-04*	-.938949E-04
TC	-.404E-07	.893E-07	.108E-06	.177425E-07
HF	-.571E-05**	.535E-05	-.197E-04	.416311E-04
FORESTP	-.490619E-07	-	-	-
TRAILS	.203004E-07***	-	-	-
PICNICTAB	-	.203004E-07	-	-
SUMNATPARK	-	.159524E-05	-	-
SWIMMIMG	-	-.734884E-05***	-	-
TENTC	-	-	-.228780E-07	-
SUMCAMPS	-	-	-.323342E-06	-
SUMWILD	-	-	-	.235090E-05
DESIGW	-	-	-	-.390465E-08

Conclusions

Insufficient data on each forest necessitate pooling of observations for forests in the same region. This encourages analysts to use weighted regression to equate the sampling distribution with the population distribution for the purpose of inference about the relevant population. However, differences in coefficient estimates of weighted and unweighted regression points to model misspecification due to pooling of different samples.This can be seen from the significant interactions of weights with the covariates included in the model. Heteroskedasticity corrected standard errors increases efficiency of the estimates but it is still biased. Therefore, it is best to include interactions of the weights with model covariate in a unweighted regression.

References

1. Barry C. Arnold and David Strauss, Pseudo likelihood Estimation: Some Examples, *The Indian Journal of Statistics*, Series B, Vol. 53, No.2, (Aug. 1991) 233-43.

2. Besag J. E., Statistical Analysis of Non-Lattice Data, *The Statistician*, Vol.24(1975) 175-195

3. Besag J. E., Efficiency of Pseudolikelihood Estimators for Simple Gaussian fields,*Biometrika*, Vol.64(1977) 616-18.

4. Bloom, D.E. and T.L. Idson, "The practical Importance of Sample Weights", Proceedings of the survey Research Methods Section.
http://www.amstat.org/sections/SRMS/Proceedings

/papers/1991_106.pdf

(Accessed December 8, 2009)

5. Cox, D.R. and N. Reid ,"A Note on Pseudo likelihood Constructed from Marginal Densities", *Biometrika*,Vol. 91, No. 3 (Sep., 2004), pp. 729-737

6. DuMouchel H. William and Greg J. Duncan, "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples", *Journal of the American Statistical Association*, Vol. 78, No. 383 (sep., 1983) 535-43.

7.Edward E. Leamer, "Specification searches ad hoc inference with non experimental data" *Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons*, Inc. 1978.

8. 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", *Rev. Econ. Statist.*77 (1995): 104-112.

9. James M. Bowker et al., "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, The University of Georgia, Department of Agricultural and Applied Economics, Athens, GA.

10. Jerald F. Lawless, Negative Binomial and Mixed poison Regression, *The Canadian Journal of Statistics*, Vol.15, No.3,(Sep. 1987), 209-25.

11. Korn, E.L. and Graubard, B.I., "Examples of Differing Weighted and Un-weighted Estimates from a Sample Survey", *The American Statistician*,

Vol. 49, No.3 (August, 1995): 291-95.

12. Manski, C.F. McFadden, D., Structural Analysis of Discrete Data and Econometric Applications. *Cambridge: The MIT Press*. 1981.

13. Shaw, W.D. " On-Site sample Regressions, Problems of Non-Negative Integers, Truncation, and Endogenous Stratification", Journal of Econometrics, Vol.37,(1998):211-223.

14. Roberto Martinez-Espineiraa and Joe Amoako-Tuffour, Recreation Demand Analysis under Truncation, Over dispersion and Endogenous Stratification: An Application to Gros Morne National Park, *Journal of Environmental Management* 88, 2008, 1320-32.

15. X. Wang et al., Asymptotic Properties of Estimators in Maximum Likelihood Estimation, *Journal of Statistical Planning and Inference* 119(2004) 37-54

16. Winship, Christopher and Larry Radbill, Sampling Weights and Regression Analysis, *Sociological Methods and Research*, Vol. 23 No. 2, Nov. 1994, 230-257

Appendix

Table 6: Summary Statistics for General Forest Area

	Mean1	Min	Max
HOTCONT	0.457	0	1
MOUNTAIN	0.190	0	1
SUBTROP	0.352	0	1
FORESTP	44.22	0.085	85.468
SUMWILDERN	.007	0.007	
INCE	21619.06	9910.434	90831.38
AGE	43.430	17.5	75
GENDER	0.160	0	1
PEOPVEH	2.254	1	10
OSITE	0.235	0	1
OVERNTE	0.225	0	1
TC	45.108	0	1221.672
HF	0.328	0	1
NFV12MO1	13.793	1	53
NOBS	1979		

Table 7: Summary Statistics for Day Used Developed Sites

	Mean1	Min	Max
HOTCONT	0.373	0	1
MOUNTAIN	0.281	0	1
SUBTROP	0.346	0	1
PICNICTAB	163.485	1	1258
SUMNATPARK	10.927	0	204
SWMMING	5.619	0	9
INCE	22808.02	8006.103	105597.6
AGE	44.063	17.5	75
GENDER	0.328	0	1
PEOPVEH	2.835	1	10
OSITE	0.325	0	1
OVERNTE	0.111	0	1
TC	64.339	0.024	1150.758
HF	0.328	0	1
NFV12MO1	8.533	1	53
NOBS	2394		

Table 8: Summary Statistics for Overnight Used Developed Sites

	Mean1	Min	Max
HOTCONT	0.374	0	1
MOUNTAIN	0.307	0	1
SUBTROP	0.319	0	1
SUMCAMPS	34.934	1	247
TENTC	452.149	22	1254
INCE	22570.38	9033.333	106902
AGE	42.693	17.5	75
GENDER	0.273	0	1
PEOPVEH	2.656	1	10
OSITE	0.331	0	1
OVERNTE	0.574	0	1
TC	42.405	0.296	728.2
HF	0.139	0	1
NFV12MO1	7.461	1	53
NOBS	1707		

Table 9: Summary Statistics for Wilderness

	Mean1	Min	Max
HOTCONT	0.412	0	1
MOUNTAIN	0.071	0	1
SUBTROP	0.517	0	1
SUMWILDERN	1.008	0	245
DESIGW	35187.1	13812	118337
INCE	26142.53	13052.6	111898.3
AGE	38.355	17.5	75
GENDER	0.276	0	1
PEOPVEH	2.754	1	9
OSITE	0.294	0	1
OVERNTE	0.297	0	1
TC	62.588	1.466	634.357
HF	0.075	0	1
NFV12MO1	5.442	1	53
NOBS	622		

Table 10: Summary Statistics: Weighted Means

	GFA	DUDS	OUDS	WILD
HOTCONT	0.334	0.300	0.506	0.252
MOUNTAIN	0.124	0.255	0.207	0.076
SUBTROP	0.542	0.444	0.287	0.671
FORESTP	44.425	-	-	-
TRAILS	301.166	-	-	-
PICNICTAB	-	152.148	-	-
SUMNATPARK	-	11.571	-	-
SWIMMING	-	5.082	-	-
SUMCAMPS	-	-	39.090	-
TENTC	-	-	565.453	-
SUMWILDERN	-	-	-	6.391
DESIGW	-	-	-	-
INCE	20893.5	22514.06	22213.99	26564.59
AGE	45.364	46.913	46.598	41.318
GENDER	0.135	0.272	0.260	.205
PEOPVEH	2.112	2.733	2.427	2.629
OSITE	0.139	0.215	0.202	0.269
OVERNTE	0.162	0.047	0.437	0.237
TC	40.059	69.535	47.9986	99.323
HF	0.372	0.165	0.170	0.143
NFV12MO1	14.638	8.483	8.008	8.537