

*Accounting for Geographic Heterogeneity in Recreation Demand  
Models*

DRAFT -Comments Welcomed

Presented paper  
AAEA Annual Meetings, July 2008  
Orlando, Florida

Kavita Sardana (presenter)  
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

C. Megan Starbuck  
Assistant Professor  
Department of Economics and International Business  
New Mexico State University, Las Cruces

Donald B.K. English  
National Forest System, Recreation and Heritage Resources  
USDA Forest Service, Washington, D.C.

*Do not cite or distribute without permission of authors.*

## *Introduction*

For land and water management and policy purposes, federal and state government agencies in the U.S. have a need for estimates of the economic value of outdoor recreation. For example, as part of national forest management and planning under the Forest and Rangeland Renewable Resources Planning Act, the USDA Forest Service periodically assesses net economic benefits (consumers surplus) of forest-based outdoor recreation (e.g., see Rosenthal and Loomis, 2001). Outdoor recreation values estimated by specific recreation activity and geographic regions are most useful to agencies such as the USDA Forest Service for informing natural resource management and planning.

In economic analysis, there are several alternative approaches for estimating outdoor recreation values by activity (e.g., hiking, boating, bicycling, camping, snow skiing) and geographic region (e.g., Northeast, Southeast, Midwest, Southwest, Mountain West, West Coast). A common approach is the full pooled data-dummy variable approach. In this approach, data are pooled across activities and regions and dummy variables and or dummy interaction variables are used to account for activity and geographic heterogeneity. Another approach is the partial pooled-dummy variable approach. In this approach, activity and geographic heterogeneity is accounted for through a combination of data pooling and dummy variables. For example, activity data may be pooled and separate models with activity dummy and/or activity dummy interaction variables are estimated by region. Activity heterogeneity is accounted for by the regressors while geographic heterogeneity is accounted for by estimating unique models for each regional data set. A third approach is the full separate data and model approach. In this approach, unique demand models are estimated for each activity by region using separate data sets.

A fourth and unique approach reported in this paper is the mixed effects modeling approach. In this approach, activity and geographic heterogeneity are accounted for through econometric techniques which capture model error term correlations. Using data from the National Visitor Use Monitoring System (NVUM), the overall purpose of this paper is to examine the feasibility of using the mixed effects approach to estimate recreation demand models using the travel cost method. The rest of this paper is organized as follows. In the next section, we provide background on data collection, NVUM and issues related to recreation value estimation including spatial heterogeneity. Next, we discuss the methodology we use to account for spatial heterogeneity

in the estimation of outdoor recreation activity values across U.S. geographic regions and report results. Some conclusions are provided in the final section.

### ***Data Collection, NVUM and Estimation Issues***

In 1998, the USDA Forest Service began developing NVUM. NVUM represents a comprehensive effort to scientifically estimate recreation visitation levels on National Forest lands on a continuous basis. After initial development and pre-testing, data collection began in 2000. In its first 4-year cycle, NVUM collected data from 120 National Forests. The preliminary or master dataset for the first cycle of on-site surveying (2000-2003 inclusive) contains 90,542 individual recreation visitor observations from 7,532 different sites aggregated from 120 National Forests and includes more than 200 variables per observation. The NVUM data provide a unique data set and opportunity for estimating National Forest recreational values by activity at the national and regional levels. Prior assessments of National Forest recreational values at the national and regional levels basically represent meta-analysis summaries of previously published value estimates (Rosenberger and Loomis, 2001).

For management purposes, the U.S. Forest Service divides the U.S. into nine regions (Figure 1). As management plans are developed for National Forests across these different regions, planners and managers often need region- or even forest-specific estimates of recreation activity values. Eventually, NVUM may provide enough data at the forest level to estimate forest-specific recreation demand models and values. However, at this point in time, data are only sufficient for estimating models and values at the regional level.

A natural question is, What causes heterogeneity in outdoor recreation values across U.S. regions? We identify two major sources of this heterogeneity: 1) differences in site characteristics across regions which influence recreation activity quality (e.g., terrain and snowfall across regions which influence downhill skiing quality); and 2) differences in user populations across regions which influence recreation visitor preferences (e.g., age and income which influence preferences for backpacking). From both planning and management and statistical estimation perspectives, it is important to properly account for these and other potential sources of spatial heterogeneity when estimating recreation activity values across regions.

### ***Estimation Methodology And Results***

With respect to recreation activities and geographic regions, the NVUM data set represents a two level grouped data set (e.g., number of visits by primary activity in each of the nine regions). Thus, in the NVUM data set, there are two nested blocking factors. Primary activity random effects are nested within region specific random effects. There may be heterogeneity from region to region and primary activities and consequently correlation for observations obtained from the same region and for the same primary activity. A statistical method which can be used to account for the expected correlation structure is a linear mixed effect model where we have random effects for the region, primary activity and primary activity within region. Linear mixed models or multilevel or hierarchical models are mixed in the sense that they allow fixed and random effects and are generalized in the sense that they are appropriate not only for continuous Gaussian response but also for binary, count and other limited dependent variable models which are frequently used to estimate travel cost recreation demand functions.

Using the above approaches and the NVUM data set, we estimate recreation demand functions and values for 14 recreation activities for the nine U.S. Forest regions. The trip quantity measure of interest is annual recreation activity visits to a National Forest. Since the visitor at site has made at least one visit to the National Forest, the distribution of annual visits is truncated at zero. We begin the analysis by modeling number of visits by using truncated Poisson distribution. Cameron and Trivedi (2005) discuss the limitation of using Poisson distribution in analyzing count data because of its assumption of equi-dispersion. In count data, variance usually exceeds the mean; thus the ability of a Poisson distribution to appropriately describe the data is limited. For this reason, in our analysis we tried modeling annual visits using a Truncated Negative Binomial distribution. Statistically, a Negative Binomial Distribution can be derived from the Poisson distribution by incorporating latent heterogeneity in the conditional mean - assuming that the random component accounting for latent heterogeneity is independent of the error term and follows a gamma distribution.

Estimating recreation demand equations and values across regions using a pooled zero truncated negative binomial count regression model would have two major drawbacks. First, according to Galwey (2006), it would ignore variation among different geographic regions. Second, it would ignore variations among observations with in each geographical region. Thus pooled regression in our case would not give consistent estimates.

Therefore, we model annual visits to a National Forest by grouping the entire data set according to geographical region. We then estimate a zero truncated Poisson model which accounts for specific heterogeneity in the mean. The reason we do not use a negative binomial distribution is that doing so would result in accounting for region specific heterogeneity twice. As mentioned above, a negative binomial distribution can be derived statistically from the Poisson by adding latent heterogeneity in the mean. Adding region specific random effects for the second time would result in double accounting of latent heterogeneity (Greene, 2007). Thus, we model annual visits to a National Forest with a truncated Poisson distribution along with region specific random effects. Region specific random effects follow a standard normal distribution. Greene calls this the Poisson mixed with log normal. The general non linear mixed effect model is given by

$$y_{ij} = f(\theta_{ij}, x_{ij}) + \epsilon_{ij} \quad (1)$$

$$i = 1, 2, 3, 4 \dots 9 \quad (2)$$

$$j = 1, 2, \dots n_i \quad (3)$$

i is the number of groups, j denotes number of individuals in group i.

$$\theta_{ij} = A_{ij}\beta + B_{ij}b_i \quad (4)$$

where  $b_i \sim N(0,1)$  In his working paper, Greene (2007) discusses the derivation of Poisson lognormal mixture. We have derived the Poisson log normal mixture model for truncated Poisson distribution, relevant for our modeling. The conditional truncated Poisson distribution with region specific heterogeneity can be derived by reparameterizing the Poisson distribution parameters. In our analysis, we call it

$h\lambda_{ij}$

$$prob(Y = y_{ij} > 0 | X_{ij}, b_i) = \frac{EXP(-h\lambda_{ij})(h\lambda_{ij})^{y_{ij}}/\Gamma(1 + y_{ij})}{1 - EXP(-h\lambda_{ij})} \quad (5)$$

where,

$$h\lambda_{ij} = EXP(\alpha + x'_{ij}\beta + b_i) \quad (6)$$

where,

$$\lambda_{ij} = EXP(\alpha + x'_{ij}\beta) \quad (7)$$

The unconditional truncated Poisson distribution with region specific heterogeneity is given by integrating out  $b_i$ ,

$$P(y_i|x_i) = \int_{-\infty}^{\infty} \frac{EXP[-EXP(b_i)\lambda_{ij}][EXP(b_i)\lambda_{ij}]^{y_{ij}}}{\Gamma(1+y_{ij})(1-EXP[-EXP(b_i)\lambda_{ij}])} \phi(b_i) db_i \quad (8)$$

where  $\phi(b_i) \sim$  gives the standard normal density function.

The conditional log likelihood for the truncated poisson distribution with normal random effects is given by,

$$\ln L(y_{ij}|\lambda_{ij}, b_i) = [y_{ij}(\alpha + x'_{ij}\beta + b_i)] - 2h\lambda - \ln\Gamma(1+y_{ij}) \quad (9)$$

The unconditional log likelihood can be estimated by integrating out  $b_i$

$$\ln L = \int_{-\infty}^{\infty} [y_{ij}(\alpha + x'_{ij}\beta + b_i)] - 2h\lambda - \ln(\Gamma(1+y_{ij})) \phi b_i db_i \quad (10)$$

where,  $\phi b_i$  is the standard normal distribution for region specific random effects.

### ***Empirical Estimation***

By maximising the unconditional log likelihood function with respect to the parameters of the model we can get the maximum likelihood estimates. According to Greene, integrals in the log likelihood function do not exist in closed form. And therefore, the quadrature based approach suggested by Butler and Moffit(1982) can be used for approximation,

$$E(Y_{ij}|X_{ij}, b_i) = h\lambda_{ij} = EXP(\alpha + x'_{ij}\beta + b_i) \quad (11)$$

The mean visits

$$E(Y_{ij}|X_{ij}, b_i) = f(\text{PEOPVEH}_{ij}, \text{GENDER1}_{ij}, \text{ONITE}_{ij}, \text{PRIMARY ACTIVITY}_{ij}, \text{TC}_{ij})$$

The independent covariates in our analysis  $x_{ij}$ , include: people traveling in a vehicle(PEOPVEH); gender of jth visitor in ith region (GENDER1: 1=female; 0=male); an indicator of staying overnight at the National Forest for jth individual from the ith region (ONITE: 1=stayed overnight on National Forest=1; 0=day user); an indicator for high frequency visitors (HF: 1 if annual visits  $\geq 15$ ; 0 otherwise); travel cost for jth individual from ith region;

and an indicator variable for the primary activity for  $j$ th individual from  $i$ th region. For our initial analysis reported in this paper, we have restricted primary activities to just 5 of the 14 primary activities in the NVUM data set. These primary activities include, HIKING, CAMPING, FISHING, HUNTING, TRAIL USE and SKIING. The travel cost variable  $TC$  is calculated as,

$$TC = 0.12 * 2 * PRACD1S + (0.333 * (INCOME / 2000)) * 2 * TIME2 \quad (12)$$

where  $PRACD1S$  is the one way distance to the national forest and a per mile cost of dollar 0.12 is used to convert travel miles to travel costs. We also include the opportunity cost of time in travel costs valued at one-third the wage rate. Individual wage rate is calculated as the annual  $INCOME$  proxy divided by 2000 hours and  $TIME2$  is the one-way distance from an individual  $s$  home to the National Forest.

## ***Results***

The results of the non-linear mixed effects regression data analysis are shown in Table 1. All parameters in the model are highly significant. The travel cost variable ( $TC$ ) coefficient is negative as expected theoretically. From an economic theory perspective, we do not have unambiguous expectations on the signs of the other parameters. The negative sign for the gender variable ( $GENDER1$ ) coefficient indicates that females have a lower demand for outdoor recreation as compared to males, a result often observed in previous recreation demand studies (Rosenberger and Loomis, 2001).

The negative sign for the number of people in a vehicle variable ( $PEOPVEH$ ) coefficient indicates that annual trips are positively correlated with group size. The negative sign on the overnight visitation variable ( $ONITE$ ) suggest overnight visitors take less annual trips as compared to day users. Because overnight are typically more expensive than day trips, the negative sign on  $ONITE$  may indicate a theoretically consistent negative price effect. The negative sign on  $PEOPVEH$  may also reflect a negative price effect if having more people in a vehicle makes trips more expensive (e.g., necessitates the use of a larger vehicle and more gasoline consumption).

The primary activity variables indicate preferences for recreation trips compared to the base of general recreation. The positive sign on the coefficients for hiking ( $HIKE$ ), fishing ( $FISH$ ), hunting ( $HUNT$ ), skiing ( $SKI$ ) and non-motorized trail use such as mountain biking ( $TRAIL$ ) indicates that these

activities are more preferred than general recreation with skiing showing the strongest positive effects. The negative sign on the coefficient for camping (CAMP) indicates that camping is less preferred than general recreation.

### ***Conclusions***

The primary objective of the analysis reported in this paper was to test the feasibility of using a mixed effects approach to estimate recreation demand models which account for geographic heterogeneity. Using data from the National Visitor Use Monitoring System (NVUM) we estimated a national-level recreation demand model using this approach. The approach generated strong parameter estimation results from both theoretical and statistical perspectives. Thus, we encourage further testing of the mixed effects approach and comparisons with other approaches for dealing with activity and geographic heterogeneity when estimating consumer demand models.

### ***References Cited***

Bowker, J.M., and V.R. Leeworthy. 1998. Accounting for Ethnicity in Recreation Demand: A Flexible Count Data Approach. *Journal of Leisure Research*. 30(1):64-78.

Cameron and Trivedi. *Microeconometrics, methods and applications*, cambridge university press.

Galwey, N.W. *Introduction to mixed modelling, beyond regression and analysis of variance*, John Wiley and Sons, Ltd.

Greene, William. *Functional Form and Heterogeneity in Models for Count Data*, Working paper, April 2007, Department Of economics, Stern school of business, NYU.

Rosenberger, R.S. and J.B. Loomis. 2001. *Benefit Transfer of Outdoor Recreation Use Values*. General Technical Report RMRS-GTR-72, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 1-59.

Vaughan, W.J., and C.S. Russell. 1982. Valuing a Fishing Day: An Application of a Systematic Varying Parameters Model *Land Economics*. 58(4):450-463.



## APPENDIX

Table 1. Nonlinear Mixed Effects Regression Results

VARIABLES	COEFFECIENT
INTERCEPT	1.590* (0.080)
GENDER1	-0.031* (0.003)
PEOPVEH	-0.032* (0.001)
ONITE	-0.138* (0.005)
HIKE	0.032* (0.005)
CAMP	-0.072* (0.007)
FISH	0.029* (0.005)
HUNT	0.028* (0.006)
TRAIL	0.047* (0.006)
SKI	0.091* (0.005)
TC	-0.003* (0.000)
HF	1.981* (0.004)
SIGMA	0.99

\*1% significance level

GCONV Convergence Criterion Satisfied.

Dual Quasi Newton, Optimization Technique Used

. Adaptive Gaussian Quadrature Integration Method Used.

Table2. U.S. Forest Service Regions

---

---

Northern Region (R1)	Pacific Southwest Region (R5)
Rocky Mountain Region (R2)	Pacific Northwest Region (R6)
Southwestern Region (R3)	Southern Region (R8)
Intermountain Region (R4)	Eastern Region (R9)
Alaska Region (R10)	

---

---

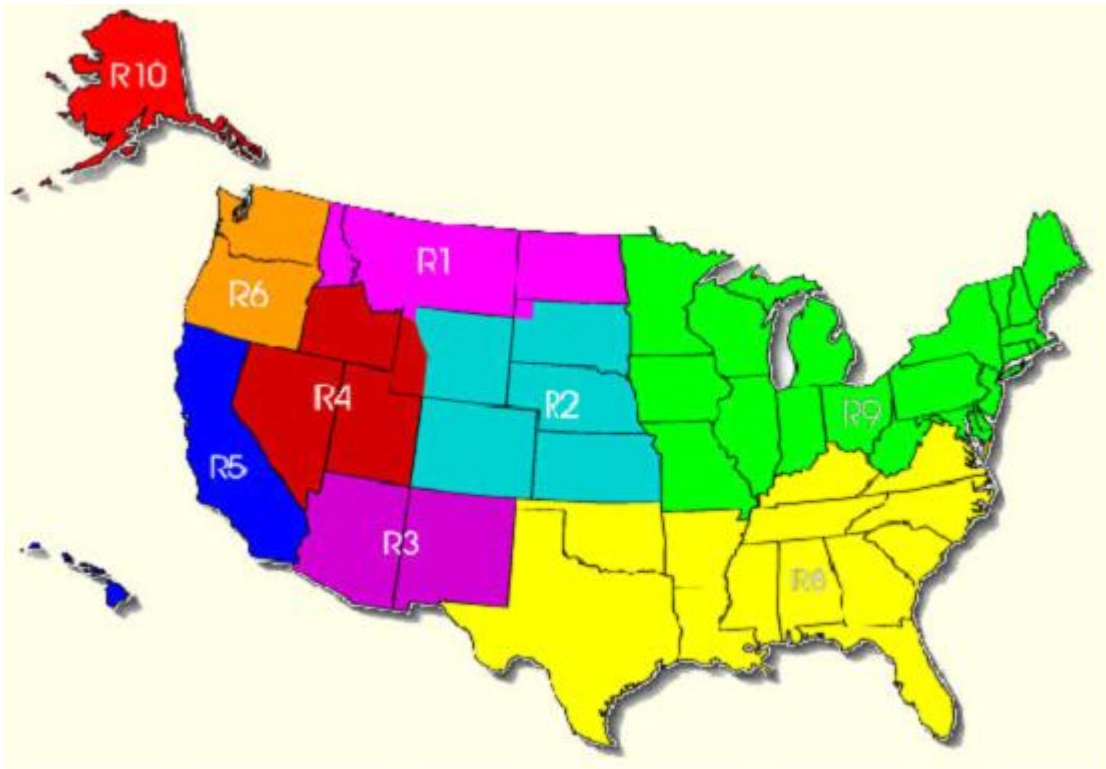


Figure 1: U.S. Forest Service Regions