Modeling Recreation Demand for British Forests

An Application of an Endogenously Stratified and Truncated Log-Normal Distribution

Jeffrey Englin* Ayele Gelan** Arwin Pang***

* Morrison School of Agribusiness and Resource Management, Arizona State University, U.S.

** Market Opportunities Theme, International Livestock Research Institute, Nairobi, Kenya

***Department of Public Finance, Feng-Chia University, Taiwan

Abstract

Evaluating the value of British forests is important for forest policies' making. Since the number of visitors to British forests is quite large and each visitor's visiting frequency is high, the conventional count data models which focus on small integers may not cope adequately with this big number of visitation. This study utilizes newly developed endogenously stratified and truncated distributions to model the recreation demand for British Forests.

Keyword: British forest recreation; count data models; endogenously stratified and truncated continuous distributions

1 Introduction

Forest recreation has been a popular pastime activity among UK households in recent decades. For example, the total number of visitors to woodlands has increased from 24 million to 50 million between early 1980s and late 1990s (Christie et al, 2005; NAO, 1986; Benson and Willis, 1992). UK day visit survey shows that the number of visitors has risen up annually by 4.3 percent from 1994 to 1998. This indicated that there has been a significant shift in public taste on forest recreation related to goods and services.

The rising trend in public demand on forest recreation would cause the need of public investment on improving quantities and qualities in recreation infrastructure at forest sites. The source of money would come from tax revenue and government's budget. In order to make wise and sustainable policy decisions, Forestry Commission has sponsored numerous valuation studies to assess the impact of forest recreation in Great Britain. (Gelan et al, 2007; Christie et al, 2006; Christie et al, 2005; Hill et al, 2003; Scarpa, 2003; Willis et al, 1989, 2000 and 2003).

However, there are considerable differences in the welfare measurement obtained by these studies. Travel cost approach has been popular among researchers for modeling forest recreation demand. The parameters of travel cost models are estimated by maximum likelihood method and the estimated parameters are used to calculate consumer surplus which is an essential input in cost-benefit analysis, environmental impact assessments and natural resource planning for public authorities. But the results crucially depend on the assumptions of distributions. Therefore it is important to test model specification before going to any policy implication.

In this study three model specifications are applied to a forest recreation survey in forty-four forests in the UK. Two are count data models which are Poisson and negative binomial while the third one is a continuous distribution, the Log-Normal distribution. Each distribution is corrected by endogenous stratification and truncation to account for the on-site sampling methodology that was employed to collect our data. A pair-wise comparison for the three alternative models is undertaken by a two-part likelihood ratio non-nested testing procedure (Vuong, 1989; Englin and Lambert, 1995), and then welfare measurements are calculated.

The remainder part of this paper is structured as follows. Section 2 describes our three theoretical models. Section 3 is survey data. Section 4 reports estimation results. Section 5 discusses the results of model estimations, the non-nested testing procedure and welfare measurement. Section 6 is the conclusion.

2 Models

Count data travel cost models have been popular for years because of the intuitive nature of these distributions. For example, Poisson and negative binomial distributions are discrete and non-negative. These two characteristics exactly feature dependent variables in the typical travel cost models.

These models are based on the linear exponential demand models. The linear exponential model can be characterized as:

$$Y = \exp(X\beta)$$

Where Y is the number of trips to a site, X is a vector of demand shifters (including travel cost and demographic characteristics of the visitors in the sample) and β is a vector of estimated coefficients. In this study, we have several different sites pooled so that site characteristics should be included in X as well. Consumer surplus per trip is simply calculated in this model as $1/\beta_{tc}$ where β_{tc} is the parameter on the travel cost variable, the demand slope coefficient (Englin and Shonkwiler, 1995).

Although Poisson and negative binomial distributions have been employed to estimate the recreation demand by their non-negative and integer attributes, for on-site sampling data, these distributions should be corrected by endogenous stratification and truncation. (Shaw, 1988; Englin and Shonkwiler, 1995).

However, these models are basically developed in the context when the number of trips is small. Recently, Englin and Nalle (2005) have proposed a class of non-negative endogenously stratified and truncated *continuous* distributions. One of them is log-normal distribution. In our case, this continuous distribution is appealing since many trips are taken to Scottish forests. Each econometric model is presented below.

Poisson distribution

The Poisson distribution corrected for both zero-truncation and endogenous stratification can be expressed as:

$$\mathrm{P}(Y_i \mid Y_i \geq 1) = \frac{e^{-\lambda}\lambda^{y-1}}{(y-1)!} \tag{1}$$

The expected value of the dependent variable equals to its variance as

$E(Y) = V(Y) = \lambda.$

Negative binomial distribution

The negative binomial probability distribution after correcting for zero-truncation and endogenous stratification can be represented as:

$$\mathrm{P}(Y_i \mid Y_i \geq 1) = \frac{y\Gamma\left[y + (1/\alpha)\right]\left\{(1/\alpha)/\left[(1/\alpha)+\mu\right]\right\}^{(1/\alpha)}\left\{\mu/\left[(1/\alpha)+\mu\right]\right\}^{y}}{\mu\Gamma(y+1)\Gamma\left(\frac{1}{\alpha}\right)} \tag{2}$$

Γ represents the gamma function, α is the size-biased over-dispersion parameter.

$E(Y) = \mu$

$V(Y) = \mu(1 + \mu\alpha)$

The difference between Poisson and Negative binomial distribution is that Negative binomial distribution relaxes the identical assumption on *Y*’s expected value and variance by an over-dispersion parameter α.

Log-normal distribution

The log-normal probability distribution corrected for endogenous stratification and zero-truncation is denoted by:

$$\mathrm{P}(Y_i \mid Y_i \geq 1) = \frac{\exp\left\{-0.5\left[\left(\ln y - \mu^*\right)/\sigma^*\right]^2\right\}}{\sigma^* \sqrt{2\pi} \exp\left(\mu^* + 0.5\sigma^{*2}\right) \Phi\left[\left(\mu^*/\sigma^*\right) + \sigma^*\right]} \qquad 3$$

Where

$\mu^* = E(\ln y)$

$\sigma^* = V(\ln y)$

The expected value of the dependent variable and the variance are given by:

$E(Y) = \exp(\mu^* + 0.5\sigma^{*2})$

$V(Y) = \exp(2\mu^* + \sigma^{*2})\left\{\exp(\sigma^{*2}) - 1\right\}$

3 Data

Data for this study was obtained from a forest visitors' survey in the UK. The survey was commissioned by the UK Forestry Commission and conducted by a team of researchers from the Macaulay Institute and University of Gloucestershire. It was carried out during July, August and September in 2002 at forty four forest sites in England, Scotland and Wales. An initial assessment of 101 forest sites was undertaken by contacting site managers and collecting information on forest attributes. From this list 44 sites were selected. The selection was based on stratifications by geographically, ownership and visitor numbers. In order to ensure that differences in expenditure patterns among the three selected countries were captured, an equal number of sites were selected in each country (Hill et al, 2003). There are thirty sites among forty four woodland sites owned by the National Forest Enterprise, seven sites owned by Woodland Trust and the rest eight sites owned by the RSPB (Royal Society for Protection of Birds). A quota of 45 interviews was set for each site, but this was not achieved in a few sites where visitors were particularly sparse.

Interviews were conducted at entrance or exit points of each forest or woodland site. Each site was assigned one interviewer, who selected respondents for interview on continuous basis which is approaching the next person after completing the previous interview. Totally 1,906 face-to-face interviews were conducted with adults over 16 years old. For group visitors, only one person was selected from a group for interview. The questionnaire was structured to collect quantitative and qualitative information on various aspects of forest recreation including visit characteristics (e.g. day trips or overnight stays, motivation for the trip, etc), expenditure on different categories (e.g. travel cost, food, etc) and socio-economic characteristics of visitors (age, sex, income, etc) and attitudes visitors had for nature conservation. Respondents were asked how often they visited the forest recreation site during the last 12 months and then they indicated the frequencies. Hill et al (2003) provides further details of the study and data.

In addition to the data collected from on-site forest visitors, the implementation of travel cost method in the context of multi-site locations needs the information about forest sites such as existence of water features in the forest area, the size of population within certain perimeters of the forest sites, the size of the forest site and trail lengths in the forest for walk. This study is benefited from an additional database which was created by Forestry Commission study (Hill et al, 2003).

Fourteen variables are included in our analysis. Table 1 displays each variable's definition and its mean value. The first two variables directly relate to travel behavior in terms of forest recreation frequency per year and distance from respondent's home to the forest site. The remaining twelve variables can be classified into three groups. The first group is the socioeconomic characteristics of respondents including age, pre-tax annual income and three variables related to respondent's education level. The second group consists of two dummy variables related to forest attributes which are coniferous forest and the existence of water feature in the forest areas. The third group is related to availability of recreational infrastructure at forest site including car park capacity in picnic area, trail length and availability of viewpoints.

4 Estimation results

This section presents results from the three models' estimation. It should be noted that there are few individuals who went to forests hundreds times each year. These people clearly have different distribution pattern from those who make moderate numbers of trips. Thus we truncate the data by the annual number of trips of twenty five. We present comparison of parameter estimation, welfare analysis and model selection tests as follows.

4.1 Parameter estimation

Table 2 presents the results of parameter estimates for the three model specifications. The constant terms in all models are positive and significantly different from zero ($p<0.05$). Similarly, the variable DISTANCE has the expected negative significantly coefficient.

The negative binomial model is preferred to the Poisson specification because the negative binomial dispersion parameter (α) is significantly different from zero ($p < 0.05$). The scale parameter for the log-normal model is significantly different from zero ($p <0.05$) which shows the log-normal model is preferred to the Poisson too. The other evidence is the Vuong test which will be discussed in the section below. After all, a comparison of the log-likelihood values of the three functional forms suggests that the log-normal specification is the model fit our data best.

4.2 Model selection

We use a non-nested model selection test to select the best model among three alternative model specifications by a two-step procedure (Vuong, 1989; Englin and Lambert, 1995). In the first step, we compare the sample variance of log likelihood ratio to the critical value from a multivariate χ^2 distribution. If the calculated value of sample variance exceeds the multivariate χ^2 value, the null hypothesis that two conditional models are distinguishable is rejected. For the rejection case, Vuong develops a second step, a directional test, to indicate either that one model dominates the other or that neither model is preferred. We discuss three pair-wise comparisons of model selection below.

Case 1. Poisson distribution and negative binomial

For step 1, the calculated variance of the LR statistic, multiplied by the number of observations, exceeds the critical value of the multivariate χ^2 so we reject that Poisson and negative binomial distributions are the same. Then we proceed to the step 2. From the directional test, the (sample-weighted) t-statistic is less than zero, it indicates a strong preference for the negative binomial model ($P < 0.01$). In addition, as we mentioned before the negative binomial is also preferred by standard log-likelihood ratio (LR) procedures since the estimate of α is significant different than zero.

Case 2. Poisson and Lognormal

In step 1, we reject that Poisson and Lognormal distributions are the same by the calculated variance of the LR statistic, multiplied by the number of observations, exceeds the critical value of the multivariate χ^2. In second step, the (sample-weighted) t-statistic is less than zero, which means that the directional test indicates that lognormal is preferred ($P < 0.01$).

Case 3. Negative binomial and log-normal

For step 1, the calculated variance of the LR statistic, multiplied by the number of observations doesn't exceed the critical value of the multivariate χ^2, so we cannot reject the null hypothesis that these distributions are indistinguishable and hence there is no need undertake the second step test.

The Vuong tests support the finding that the negative binomial and lognormal are preferred over Poisson.

4.3 Welfare analysis

An important objective of forest recreation demand analysis is to calculate consumer surplus and measure willingness to pay for consuming public goods such as forest recreation. The calculation of consumer surplus for an individual who takes a trip to a forest site would be obtained as an inverse of the coefficient of travel cost variable. In this study we use one way distance to represent individual's travel cost in the regression. So we convert the estimated parameter on distance into a travel cost by using an average travel cost per mile (cpm) and we consider a return trip as well (i.e. round-trip distance), then consumer surplus for an individual forest visitor is calculated by following formula:

$$CS = \frac{1}{(d*2)/cpm},$$

Where *d* is the coefficient of distance in the travel cost model.

Travel cost per mile in the UK was given as £0.367 pence (Inland Revenue, 2007)[1]. By applying the above formula, consumer surplus for the three alternative travel cost models is calculated. The Poisson specification yields consumer surplus of £9.08 per visit, which is the lowest amount compared to the other two functional forms. Per visit consumer surplus is given as £14.12 for the negative binomial and £15.68 for the lognormal specifications. It is useful to note that the calculated consumer surplus for the lognormal functional form which is the preferred model has the highest value.

Finally, one can also use the estimated parameters to calculate the value of different features of the forests. In this analysis the value of six forest features are calculated by

$$CS_i = \frac{1}{(d*2)/cpm}(\lambda^1 - \lambda^0)$$

The subscript *i* refers forest feature including forest cover, car parking capacity, picnic faculties, trail length, viewpoints and water features. λ^1 is the expected trips with that

[1] The New statutory mileage rates (Rev BN 2, 7 March 2001) provides per mile rates by each car type (up to 1500cc and 1500cc to 2000 cc) and by average annual mileages (up to 4000 and 4000 plus). The rates given were 40p and 45p per mile respectively up to 4000 miles for each vehicle type and 25p for additional mileages over 4000 miles for both vehicle categories. Rates for vehicles types with more than 2000cc were excluded. Accordingly, the average cost per mile was calculated as a simple average of £0.45, £0.40 and £0.25. This gives £0.367.

certain forest feature increasing one unit. λ^0 is the expected trips in the model. Table 3 shows the values for the six attributes. It is interesting to note that the welfare measures of the log-normal and Poisson are fairly closely while the negative binomial estimates are markedly different. The welfare results suggest that people prefer deciduous forests with long trails, a place to picnic and water features such as ponds, creeks or rivers. The only difference between the discrete and continuous models is the value of car parking capacity. Additional car parking capacity is a bad in both count models while it is a good in the log–normal model.

5 Conclusion

There has been a growing demand for forest recreation among British households. The high forest visitations make that existing count data models not adequately fit British forest recreation data. Model selection tests are appropriate to decide which model fit our data best. Using British Households' forest recreation survey data, this study estimate the forest recreation demand by two conventional count data models which are Poisson distribution and negative binomial, and a newly developed Log-normal distribution model. The model selection tests show that the log-normal distribution is preferred to the conventional models. It is also supported by the comparison of the log-likelihood values. Importantly, the log-normal specification yields larger consumer surplus per person per visit compared to the conventional count data models. This suggests that model selection and application would be helpful for benefit-cost analysis and natural resource management policies.

6 References

Benson, J. F. and K. G. Willis (1992) Valuing informal recreation on the forestry commission estate. *Forestry Commission Bulletin*, *104*, HMSO, London, UK.

Christie, M., N. Hanley, T. Hyde, B. Garrod, K. Swales, N. Lyons, I. Keirle1, and A. Bergmann (2005) Valuing Forest Recreation Activities: Phase 1. Report to the Forestry Commission.

Christie, M., N. Hanley, B. Garrod, T. Hyde, N. Lyons, A. Bergmann, and S. Hynes (2006) Valuing Forest Recreation Activities: Final Phase 2. Report to the Forestry Commission.

Englin, J. and D. Lambert (1995) "Measuring Angling Quality in Count Data Models of Recreational Fishing: A Non-Nested Test of Three Approaches" *Environmental and Resource Economics* 6: 389-399.

Englin, J. and L. Nalle (2005) "Continuous Endogenously Stratified and Truncated Models" Paper presented at the Methodological Developments in Revealed Preference Modeling for Non-Market Valuation session at the American Agricultural Economics Meeting in Providence, Rhode Island. July 2005.

Englin, J. and J.S. Shonkwiler (1995) "Estimating Social Welfare Using Count Data Models: An Application to Long-run Recreation Demand under Conditions of Endogenous Stratification and Truncation" *Review of Economic and Statistics* 77(1): 104-122.

Gelan, A., W. Slee, K. Brown, and P. Shannon (2007) Examining the Commercial Potential of Recreation on the National Forest Estate in Scotland, Phase 1 Report. Report to Forestry Commission Scotland.

Hill, G., P. Courtney, R. Burton, J. Potts, P. Shannon, N. Hanley, C. Spash, J. DeGroote, D. MacMillan and A. Gelan (2003) Forests' Role in Tourism, Report for the Forestry Group (Economics & Statistics) of the Forestry Commission

NAO (National Audit Office) (1986) Review of Forestry Commission Objectives and Achievements. Report by the Comptroller and Auditor General, National Audit Office, HMSO, London.

Scarpa, R. (2003) The Recreational Value of Woodlands. Forestry Commission: Edinburgh.

Shaw, D. (1988) "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation and Endogenous Stratification" *Journal of Econometrics* 37: 211-223.

Vuong, Q. (1989) "Likelihood ratio tests for model selection and nonnested hypotheses" *Econometrica* **57**: 307–333

Willis, K.G. and J.F. Benson (1989) Values of user benefits of forest recreation: some further site surveys. Report to the Forestry Commission, Department of Town and County Planning. University of Newcastle upon Tyne.

Willis, K., G. Garrod, R. Scarpa, D. Macmillan and I. Bateman (2000). Non-Market Benefits of Forestry. Report to the Forestry Commission.

Willis, K. G., G. Garrod, R. Scarpa, N. Powe, A. Lovett, I. J. Bateman, N. Hanley and D. C. Macmillan (2003) The social and environmental benefits of forests. Report to the Forestry Commission.

Table 1 – Variable Definitions

Variable	Mean	Definition
Trips	5.18 (4.98)	Number of visits in year 2001
DISTANCE	22.1 (21.5)	Distance from residence to the forest site (miles)
AGE	43.7 (14)	Respondent Age (years)
INCOME	2.78 (1.71)	Pre-tax annual household income (10 thousands)
EDU_1STD	0.232 (0.423)	Dummy variable for education for first degree
EDU_POST	0.077 (0.268)	Dummy variable for education for postgraduate qualification
EDU_OLE	0.606 (0.49)	Dummy variable for education for O' Levels GCSE's standard Grade
EDU_OTHER	0.056 (0.232)	Dummy variable for other education level
CONIFEROUS FOREST	0.402 (0.491)	Dummy variable for coniferous forest
WATER FEATURE	0.419 (0.494)	Dummy variable for water
CARCAP	1.64 (3.49)	Car park capacity (hundreds)
PICNIC	0.703 (0.458)	Dummy variable for picnic site
TRAILLEN	5.17 (2.87)	Maximum length for walking trails (km)
VIEWPOINT	0.423 (0.495)	Dummy variable for viewpoint

Table 2: Parameter estimates

	Poisson			Negative Binomial			Lognormal		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
CONSTANT	1.5100	0.1760	*	0.7510	0.3640	*	1.0400	0.2890	*
DISTANCE	-0.0202	0.0026	*	-0.0130	0.0032	*	-0.0117	0.0032	*
AGE	0.0098	0.0025	*	0.0078	0.0047	**	0.0040	0.0040	
INCOME	-0.0310	0.0206		-0.0405	0.0367		-0.0219	0.0333	
EDU_1STD	0.1650	0.0819	*	0.0836	0.1510		0.0788	0.1340	
EDU_POST	-0.2110	0.1350		-0.2380	0.2370		-0.2290	0.2100	
EDU_OLE	0.4090	0.0744	*	0.3360	0.1360	*	0.2470	0.1220	*
EDU_OTHER	0.0668	0.1540		0.3210	0.2850		0.2490	0.2370	
CONIFEROUS FOREST	0.1590	0.0888	**	0.0581	0.1580		0.1670	0.1490	
WATER FEATURE	-0.3210	0.0806	*	-0.4020	0.1450	*	-0.2730	0.1240	*
PARKING CAPACITY	0.0119	0.0109		0.0114	0.0221		-0.0014	0.0187	
PICNIC	-0.2480	0.0840	*	-0.2120	0.1580		-0.1980	0.1420	
TRAIN LENGTH	-0.0457	0.0124	*	-0.0322	0.0220		-0.0329	0.0200	
VIEWPOINT	0.1660	0.0709	*	0.1530	0.1300		0.0569	0.1130	
α				1.1800	0.3250	*			
σ							0.7250	0.0430	*
No. of observations	246			246			246		
Log Likelihood	-747			-592			-564		
Consumer Surplus per trip	9.08			14.12			15.68		

Significant at: * 5% level, ** 10% level

Table 3. Consumer Surplus Estimates for Forest Attributes*

	Poisson	*Negative Binomial*	*Log-Normal*
Coniferous Forest	-6.54 (3.44)	-1.59 (0.69)	-8.15 (2.84)
Car Capacity	-0.45 (0.23)	-0.30 (1.33)	0.06 (0.02)
Picnic	8.34 (4.39)	5.10 (2.23)	8.07 (2.81)
Trail Length	1.70 (0.89)	0.84 (0.36)	1.45 (0.50)
Viewpoint	-6.84 (3.61)	-4.40 (1.92)	-2.63 (0.91)
Water Feature	10.40 (5.49)	8.82 (3.85)	10.70 (3.73)

* Standard errors are reported in parentheses