Estimating Travel Cost Model: Spatial Approach †

Seung Gyu Kim[1*]

J. M. Bowker[2*]

Seong-Hoon Cho[1*]

Dayton M. Lambert[1]

Donald B.K. English[3]

C. Meghan Starbuck[4]

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

Corresponding author: S. Cho. University of Tennessee, 2621 Morgan Circle, 314-A Morgan Hall, Knoxville, TN 37996-4518, Tel: +1 865 974 7408. Fax: +1 865 974 9492, e-mail: scho9@utk.edu.

† Incomplete draft. Work in progress.

[1] Department of Agricultural Economics, University of Tennessee, Knoxville, TN.

[2] Southern Research Station, USDA Forest Service, Athens, GA.

[3] National Forest System, Recreation and Heritage Resources, USDA Forest Service, Washington, DC.

[4] Department of Economics and International Business, New Mexico State University, Las Cruces, NM.

* Senior authorship is shared by the first three authors.

Estimating Travel Cost Model: Spatial Approach

High gasoline prices made headlines in 2008 and 2009. The average gasoline price in the United States peaked at $4.05 per gallon in the second week of July, and it remained fairly high most of the third quarter of 2008, dropped significantly through September to December in 2008, and began to rise again in January 2009 (Energy Information Administration, 2009).

A survey conducted during June 9–12, 2008 while gasoline prices were rising significantly, shows that 11% of Americans were limiting or cutting back on their travel or vacations due to the rising gasoline prices (Newport, 2009). As the survey results suggest, the effect of rising gasoline prices on consumers' leisure travel patterns is assumed to be negative due to decreased wealth and disposable income and increased cost (or price) of their travel given a downward sloping demand curve for recreation, if recreation is a normal good. Nonetheless a quantitative estimate of the impact of gasoline prices on leisure travel patterns has not been explicitly explored (Englin, et al., 2003, Heberling and Templeton, 2009, Hesseln, et al., 2003, Martinez-Espineira and Amoako-Tuffour, 2008). The lack of research on the impact of gasoline price on leisure travel patterns may be due to relatively stable gasoline prices during the 80's and 90's. The impact of gasoline price on leisure travel patterns has become an important issue because consumer sensitivity to fluctuations in gasoline prices affects travel decisions (Walsh, et al., 2004) and spikes in gas prices can be nontrivial.

The travel cost model is often used to measure the benefits provided by access to public recreation sites, e.g., national parks and national forests, which have relatively minor, if any, entrance fees (Oh, et al., 2005). Hotelling (1947) is credited with the initial development of the travel cost model. Using the travel cost model, observed travelers' net economic benefit, or consumer's surplus, from visiting a recreation site is calculated as the value of access to the recreation site less the travel cost and necessary entrance fees (Heberling and Templeton, 2009). The model assumes that people travel to a recreation site if the marginal value of accessing the site is at least as large as the marginal cost of traveling to the site. The estimated consumer surplus is often used as a monetary measure of consumer welfare. The aggregate net economic benefit of access to a recreation site is estimated by aggregating average individual consumer surplus per visit over all visits.

Early travel cost studies employed the zonal travel cost model (ZTCM). The ZTCM is estimated using aggregated visitation rates and average trip costs from various geographic origin zones (Willis and Garrod, 1991). In US-based studies these zones are often counties (English and Bowker, 1996, Hellerstein, 1991). The individual travel cost model (ITCM) is preferred to the ZTCM in more recent literature because it accommodates individual visitor's inherent variation in socio-economic characteristics, and more individual data are available, e.g., National Forest Visitor Use Survey (NVUM). Since the recreation demand is estimated by number of individual visits, heterogeneity in the population that is neglected by the ZTCM is accommodated in the ITCM model. The ITCM also avoids arbitrary zone definitions required in the ZTCM. As a result, the ITCM gains better statistical efficiency than the ZTCM (Bowker and Leeworthy, 1998). Finally, the ITCM is better suited to provide inferences about individual consumer behavior.

Despite the advantages over the ZTCM, there are some modeling challenges associated with the ITCM. The first obstacle is the number of visits is positive integers, as survey respondents in on-site sampling must report at least one visit. The estimated parameters of a travel cost model are biased and inconsistent unless the truncation of zero and negative values for the number of visits is addressed properly. Recreation demand studies using a single site demand equations typically apply truncated and censored Poisson and negative binomial regression models (Creel and Loomis, 1990, Grogger and Carson, 1991, Gurmu, 1991, Hellerstein and Mendelsohn, 1993, Shaw, 1988, Siderelis and Gustke, 2000). Zero-truncated regression corrects a selectivity bias due to the on-site sampling, providing a more accurate (unbiased) estimate of the regression coefficient (Bubnova, 2000).

The second challenge is that on-site surveys are endogenously stratified, meaning that frequent visitors are more likely to enter the sample (Shaw, 1988). Econometric estimators designed for random samples may be inconsistent or inefficient when applied to these samples (Cameron and Trivedi, 1998). This choice-based sampling was first addressed by Shaw (1988), while Englin and Shonkwiler (Englin and Shonkwiler, 1995) extended Shaw’s analysis with an application of the truncated and endogenously stratified negative binomial model. The presence of endogenous stratification can be mitigated by weighting each observation by the multiplicative inverse of number of visits (Kriesel, et al., 2005).

Another challenging issue regarding to travel cost model are spatial considerations (Kerkvliet and Nowell, 1999). For example, the spatial limits of the ITCM were identified by Smith and Kopp (1980). Bell and Leeworthy (1990) provided the reasons why long-distance tourists’ recreational decision-making is different from that of those traveling relatively short distances. As a remedy to address the issue of spatial limits, Hellerstein (1991) offered some precedent for deleting all trips of more than 1,000 miles, claiming they are multipurpose, which violates the travel cost model’s necessary assumption of weak complementarity. However, spatial considerations have not been explored extensively with regard to potential bias or inefficiency due to spatial error dependence.

This study is designed to estimate individual demand for access to a national forest. To more accurately measure recreation demand, we apply spatial heteroskedastic autocorrelation consistent (HAC) covariance estimators to deal with the spatial error dependence while we accommodate the abovementioned issues of ITCM. The spatial dependence due to the interaction between unobserved factors is addressed by incorporating spatial heteroskedastic autocorrelation consistent (HAC) covariance estimators. Unobserved factors may include different alternative recreation sites and different information about travel opportunities possibly influencing the individual's decision about traveling

As a case study, this travel cost model is applied to data collected from the Allegheny National Forest (ANF) Visitor Use Survey through the National Visitor Use Monitoring program (NVUM, 2003). Using the parameters estimated from the model, *ex ante* simulations generate forecasts of the number of visits at the status quo gasoline price compared to those in hypothetical gasoline price scenarios. The predicted number of visits from the *ex ante* simulations are used to conduct welfare analysis to examine the change of consumer surplus associated with a change in quantity of visits demanded caused by a noticeable change in the gasoline price.

Empirical Model

Model Specification

The single-site demand function using the individual travel cost method (ITCM) is specified as:

$$Y_i = f(T_i, F_i, C_i, S_i, H_i, O_i, N_i), \tag{1}$$

where Y_i is the number of visits during the past 12 months for a individual or group *i*, T_i is individual *i*'s travel cost to visit the ANF, F_i is the forest site type where individual *i* was interviewed. In addition to the travel cost and characteristics of the forest site type, that are commonly used in single-site demand functions (e.g., Shaw and Jakus, 1996), characteristics of individual *i*'s members (C_i) (i.e., number of companying children under 16 year-old and total number of people of individual *i*'s vehicle) and a dummy variable indicating survey year (i.e., S_i = 1 if surveyed in 2005, S_i = 0 if surveyed in 2001) are included to differentiate individual *i*'s characteristics and its temporal differences. An additional variable (H_i) is included to accommodate the differences between high and low frequency users (i.e., H_i =1 if number of annual visits was greater than 15, 0 otherwise) following Bowker et al. (2005).

In addition to the variables constructed based on the information from NVUM, one variables are added. Spatial heterogeneity can be caused by spatial heteroscedasticity due to the omitted variables (Conway, et al., 2008). One possible remedy to control the spatial heterogeneity is to include variables representing characteristics of the visitors' area of origin (O_i) (Dubin, 1988). O_i is unemployment rate. The inclusion of the socio-economic characteristic of the individual *i*' areas of origins in the equation is to control the spatial heterogeneity that may be caused by the omitted variables due to insufficient information from NVUM.

Model Estimation

Poisson and negative binomial regression were applied to estimate the individual travel cost model, and a negative binomial model is appropriate where over-dispersion exists in the sample data (i.e., a situation where the variance of the response variable exceeds the mean). Because on-site surveys only capture a subset of the entire population of potential visitors, the distribution of the number of visits is limited to positive integers. If zero truncation is not accommodated in the model, the estimated parameters will be biased and inconsistent, which will generate overstated consumer surplus estimates (Creel and Loomis, 1990, Englin and Shonkwiler, 1995, Grogger and Carson, 1991, Shaw, 1988, Yen and Adamowicz, 1993). The density of the negative binomial distribution truncated at zero for the count (Y_i) is defined as:

$$\Pr(Y = Y_i \mid Y > 0) = \frac{\Gamma(Y_i + 1/\alpha)}{\Gamma(Y_i + 1)\Gamma(1/\alpha)} (\alpha\mu)^{Y_i} (1+\alpha\mu)^{-(Y_i + 1/\alpha)} \left(\frac{1}{1-(1+\alpha\mu)^{-1/\alpha}} \right) , \qquad (2)$$

where μ is the intensity or rate parameter. To accommodate endogenous stratification, endemic in on-site sampling, Shaw (1988) showed that the probability of visitors being included in the sample is proportional to the number of visits taken. Thus, the inverse of the number of individual visits ($1/Y_i$) is applied to all the variables in the regression as the choice-based sample weight.

In order to incorporate spatial heteroskedastic autocorrelation consistent (HAC) covariance estimators, additional assumptions outlined by Kelejian and Prucha (2007) are necessary. Given a consistent estimator for equation (2), the disturbance vector (ε_{ij}) is assumed to be $\varepsilon_{ij} = Y_{ij} - E\left(Y_{ij}\right) = \mathbf{r}_i u$, where $\mathbf{r}_j$ is the i^{th} row of an n×n nonstochastic matrix (**R**) with unknown elements whose row and column sums are uniformly bounded in absolute value and u is an n×1 independent and identically distributed vector of disturbances with zero mean and σ^2 (Lambert and McNamara, 2009) The consistency of the asymptotic distribution of the nonstochastic determinants was proved to be consistent by Kelejian and Prucha (2007). The Epanechnikov kernel is used to adjust for covariance between cross sectional units.

In addition to quantifying factors describing demand and determining price response, the most typical reason for estimating the single site travel cost model is to calculate consumer surplus, a welfare measure commonly used in benefit-cost analysis (Ward and Loomis, 1986). Following Englin and Shonkwiler (1995), Siderelis (2001), and Heberling and Templeton (2009), the average individual consumer surplus per ANF visit (CS_i) is calculated as:

$$CS_i = -1/\beta_{travel\,cost} \ , \tag{3}$$

where, $-1/\beta_{travel\ cost}$ is the reciprocal of the travel cost coefficient. However, as groups (primarily cars) were the sampling unit as opposed to individuals, the average individual group's consumer surplus must be divided by the number of people in the group N_i to yield average individual consumer surplus per visit. Aggregate consumer surplus or the total net economic value of annual access to the ANF can then be measured by multiplying total number of estimated individual visits by the average consumer surplus per individual visit.

Estimated aggregate consumer surplus can be used to evaluate the total economic welfare associated with access to the ANF. However, the model can also be used *ex ante* to assess welfare changes associated with policy or exogenous changes to any of the descriptors in the model. For example, although future visitation is unavailable, price elasticity from the estimated travel cost demand function can be combined with existing visitation and changes in travel cost to quantify the visitation and welfare effects of higher gasoline prices. We demonstrate these effects using three different gasoline price scenarios, i.e., 25%, 50%, and 100% increase from the status quo gasoline price.

The Akaike Information Criterion (AIC) was chosen to test the performance of the model estimated with the variables that accommodate the spatial limits (the variables representing characteristics of the visitors' area of origin, i.e., unemployment rate, percentage of age group over 65, and average travel time to work at the county level, and the number of visits from the neighbor of individual's area of origin at the zip code level) relative to the model without the variables.

Study Area and Data

This study uses three primary data sets: National Visitor Use Monitoring (NVUM), census data at the county level, and environmental data using geographic information systems (GIS). The research study area is the Allegheny National Forest, located in northwestern Pennsylvania (see Figure 1). The ANF consists of 513,000 acres, and includes over 600 campsites, and venues for outdoor activities such as hiking, hunting, snowmobiling, and riding all-terrain vehicles.

Information about visitors to the ANF was collected from NVUM, which is designed to provide information on National Forests and Grasslands visitor satisfaction and use (USDA Forest Service, 2009). This survey has been conducted annually for 120 National Forests (or combinations thereof) since 2000. This on-site survey data includes reported visits during the past 12 months, forest site type, various demographic variables, and the respondent location of residence identified by zip code.

NVUM uses the double sampling technique, which was developed to measure recreational use on national forest using two steps of sampling (James, 1967). In the first step of the sampling, the days and locations that represent when and where recreational visitors can be counted are randomly selected from a stratified set of site-days. In the second step, the individual with the most recent birthday among the people in a vehicle who were last-exiting the selected recreation site is interviewed. For each site-day selected in the sample plan, on-site interviews were conducted during one of the two randomly selected 6-hour interview periods (English, et al., 2001).

The travel distances between the visitors' origins by zip code and the Allegheny National Forest were measured between their centroids using the ArcGIS tool of Network Analysis. Following Heberling and Templeton (2009) and Karp et al. (2000), round-trip travel distances (in miles) are multiplied by $0.14/mile, which is the reimbursement rate for charitable organizations specified by the Internal Revenue Service to calculate round–trip travel cost (IRS, 2004).[1]

There are four different forest site-type strata in our data. Day-use developed sites (DUDS) include picnic sites, fish viewing sites, fishing sites, interpretive sites, observation sites, playground-park sport sites, ski areas, some wildlife viewing sites, caves, visitor centers, museums, and swimming areas. Overnight-use developed sites (OUDS) include campgrounds, fire lookouts and cabins, hotels, lodges, and resorts, horse camps, organization sites, and any other overnight developed sites within Forest Service jurisdiction, whether managed by the agency or by a concessionaire. Wilderness (WILD) includes lands and waters that are part of the National Wilderness Preservation System. General forest area (GFA) includes all of the residual parts of a national forest not included in DUDS, OUDS, or WILD categories. Generally, sample points for general forest area are at trailheads. A dummy variable indicating type of forest site at which the respondent was interviewed (1 if interviewed at general forest site, 0 otherwise) was added to the model.

Consistent with the NVUM rotation across all National Forests, data for the Allegheny National Forest were collected in 2001 and 2005. Randomly selected survey dates for 2001 and 2005 were respectively 87 and 92 days throughout 12 months for each year. Hence, a dummy variable indicating survey year (1 if surveyed in 2005, 0 if survey in 2001) was used to capture

---

[1] Internal Revenue Service announces standard mileage rates annually, and change of gasoline price is a significant factor among fixed and variable costs. Alternatively, the travel cost can be calculated based on the average operating cost ($0.118/mile for 2001 and $0.141/mile for 2005) reported by the American Automobile Association (AAA).

the differences between two periods. In addition, other socio–economic variables such as unemployment rate, population distribution by age, and average travel time to work that characterize the demography of the visitors' origins were collected from the 2000 U.S. census long-form dataset at the county level because of unavailability of these variables at the zip code level. Respondent origins included 425 different zip codes and 104 counties. Although the time of the census and the NVUM did not match, given the timing of census taking, the 2000 Census data were used as proxies in the model.

Census variables included unemployment rate, age, and travel time to work. Unemployment rate is used as a demographic variable reflecting economic status of the areas of origins. Gum and Martin (1977) argued that age was the crucial variable in terms of recreation decisions, so a variable for percentage of age of 65 or older was included to examine age-wise preference for recreational demand for the forest site. Travel time to work at the visitor's area of origin was used at the county level in the model to capture the differences in urbanization, population density, and size of counties because travel time to work of a county is differentiated by these three characteristics.

References

Allegheny National forest (2009) Location information, US Forest Service.

Bell, F.W., and V.R. Leeworthy. 1990. "Recreational demand by tourists for saltwater beach days." *Journal of Environmental Economics and Management* 18(3):189-205.

Bowker, J.M., D. English, J.C. Bergstrom, and C.M. Starbuck. 2005. "Valuing national forest recreation access: Using a stratified on-site sample to generate values across activities for a nationally pooled sample." Paper presented at Selected Paper Prepared for Presentation at the American Agricultural Economics Association Annual Meeting, Providence Rhode Island, July 24-27.

Bowker, J.M., and V.R. Leeworthy. 1998. "Accounting for ethinity in recreaion demand: A flexible count data approach." *Journal of leisure research* 30(1):64-78.

Bubnova, N.B. "Governance impact on private investment: Evidence from the international patterns of infrastructure bond risk pricing." World Bank - Technical Papers.

Cameron, A.C., and P.K. Trivedi. 1998. *Regression analysis of count data*: Cambridge University Press.

Conway, D., C. Li, J. Wolch, C. Kahle, and M. Jerrett. 2008. "A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values." *The Journal of Real Estate Finance and Economics*:DOI 10.1007/s11146-11008-19159-11146.

Creel, M., and J.B. Loomis. 1990. "Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in california." *American Journal of Agricultural Economics* 72:434-441.

Dubin, R.A. 1988. "Estimation of regression coefficients in the presence of spatially autocorrelated error terms." *The Review of Economics and Statistics* 70(3):466-474.

Energy Information Administration (2009) Office energy statistics from the united states government. U.S. Weekly retail gasoline price.

Englin, J., T. Holmes, and E. Sills. 2003. "Estimating forest recreation demand using count data models." *Forests in a market economy*:341–359.

Englin, J., and J.S. Shonkwiler. 1995. "Modeling recreation demand in the presence of unobservable travel costs: Toward a travel price model." *Journal of Environmental Economics and Management* 29(3):368-377.

English, D., and J. Bowker. 1996. "Sensitivity of whitewater rafting consumers' surplus to pecuniary travel cost specifications." *Journal of Environmental Management* 47(1):79-91.

English, D.B.K., S.M. Kocis, S.J. Zarnoch, and J.R. Arnold. "Forest service natioanl visitor use monitoring process: Research method documentation." USDA National Forest.

Grogger, J.T., and R.T. Carson. 1991. "Models for truncated counts." *Journal of Applied Econometrics* 6(3):225-238.

Gum, R.L., and W.E. Martin. 1977. "Structure of demand for outdoor recreation." *Land Economics* 53(1):43-55.

Gurmu, S. 1991. "Tests for detecting overdispersion in the positive poisson regression model." *Journal of Business and Economic Statistics* 9:215-222.

Heberling, M., and J. Templeton. 2009. "Estimating the economic value of national parks with count data models using on-site, secondary data: The case of the great sand dunes national park and preserve." *Environmental Management* 43(4):619-627.

Hellerstein, D., and R. Mendelsohn. 1993. "A theoretical foundation for count data models." *American Journal of Agricultural Economics* 75(3):604-611.

Hellerstein, D.M. 1991. "Using count data models in travel cost analysis with aggregate data." *American Journal of Agricultural Economics* 73(3):860-866.

Hesseln, H., J.B. Loomis, A. González-Cabán, and S. Alexander. 2003. "Wildfire effects on hiking and biking demand in new mexico: A travel cost study." *Journal of Environmental Management* 69(4):359-368.

Hotelling, H. 1947. *Letter cited in the economics of public recreation: An economic study of the monetary evaluation of recreation in the national parks*. Washington, D.C.: U.S. National Park Service.

IRS (2004) 2005 standard mileage rates set, Internal Revenue Service, United States Department of the Treasury.

James, G.A. "Recreation use estimation on forest service lands in the united states." USDA-Forest Service, Southeastern Forest Experiment Station.

Karp, W.B., R.K. Grigsby, M. McSwiggan-Hardin, S. Pursley-Crotteau, L.N. Adams, W. Bell, M.E. Stachura, and W.P. Kanto. 2000. "Use of telemedicine for children with special health care needs." *Pediatrics* 105(4):843.

Kelejian, H., and I. Prucha. 2007. "Hac estimation in a spatial framework." *Journal of Econometrics* 140(1):131-154.

Kerkvliet, J., and C. Nowell. 1999. "Heterogeneous visitors and the spatial limits of the travel cost model." *Journal of leisure research* 31(4).

Kriesel, W., C. Landry, and A. Keeler. 2005. "Coastal erosion management from a community economics perspective: The feasibility and efficiency of user fees." *Journal of Agricultural and Applied Economics* 37(2):451.

Lambert, D.M., and K.T. McNamara. 2009. "Location determinants of food manufacturers in the united states, 2000-2004: Are nonmetropolitan counties competitive?" *Agricultural Economics* 40(6):617-630.

Martinez-Espineira, R., and J. Amoako-Tuffour. 2008. "Recreation demand analysis under truncation, overdispersion, and endogenous stratification: An application to gros morne national park." *Journal of Environmental Management* 88(4):1320-1332.

Newport, F. (2009) Gas prices having a ripple effect in americans' lives americans report that their personal financial lives are being affectedlives.Aspx.

NVUM (2003) National forest visitor use monitoring program, national project results, january 2000 through september 2005

Oh, C.-O., R. Ditton, D. Anderson, D. Scott, and J. Stoll. 2005. "Understanding differences in nonmarket valuation by angler specialization level." *Leisure Sciences* 27(3):263-277.

Shaw, D. 1988. "On-site sample regression: Problems of non-negative integers, truncation, and endogenous stratification." *Journal of Econometrics* 37:211-223.

Shaw, W.D., and P. Jakus. 1996. "Travel cost models of the demand for rock climbing." *Agricultural and Resource Economics Review*:133-142.

Siderelis, C. 2001. "Incidental trips and aquarium benefits." *Leisure Sciences* 23(3):193-199.

Siderelis, C., and L. Gustke. 2000. "Influence of on-site choices on recreation demand." *Leisure Sciences* 22(2):123-132.

Smith, V., and R. Kopp. 1980. "The spatial limits of the travel cost recreational demand model." *Land Economics* 56(1):64-72.

USDA Forest Service (2009) Recreation, heritage & wilderness programs.

Walsh, K., C.A. Enz, and L. Canina. 2004. "The impact of gasoline price fluctuations on lodging demand for us brand hotels." *International Journal of Hospitality Management* 23(5):505-521.

Ward, F.A., and J.B. Loomis. 1986. "The travel cost demand model as an environmental policy assess-ment tool: A review of literature." *Western Journal of Agricultural Economics* 11:164-178.

Willis, K.G., and G.D. Garrod. 1991. "An individual travel-cost method of evaluating forest recreation." *Journal of Agricultural Economics* 42(1):33-42.

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

Table 1. Variable names, definitions, and descriptive statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Dependent Variable					
Number of visit	Number of visits during the past 12 months for a last exiting individual group in a vehicle	21.31	48.00	1	365
Variables collected from NVUM					
Travel cost	Travel cost for round trip ($)	24.53	23.39	1.51	237.74
General forest area	Type of site at which the respondent was interviewed (1 if interviewed at general forest site, 0 otherwise)	0.29	0.45	0	1
Child	Number of accompanying children under 16 year–old	0.70	1.17	0	10
Accompanying people	Number of accompanying people in the same vehicle	2.75	1.43	1	9
Round	Dummy variable indicating survey year (1 if surveyed in 2005, 0 if surveyed in 2001)	0.58	0.49	0	1
Variables to accommodate spatial heterogeneity					
Unemployment rate	Unemployment rate of the individual *i*'s area of origin	0.04	0.01	0.02	0.06

Figure 1. The location of Allegheny National Forest

Source: Allegheny National forest, 2009