Kuhn-Tucker Estimation of Recreation Demand –
A Study of Temporal Stability

Subhra Bhattacharjee, Catherine L. Kling and Joseph A. Herriges
Iowa State University
Contact: subhra@iastate.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26-29, 2009

Kuhn-Tucker Estimation of Recreation Demand – A Study of Temporal Stability

Recreation demand models have been extensively studied in the environment economics literature, because they involve valuation of environmental goods and they allow the estimation of changes in welfare from changes in the environmental quality of outdoor recreation sites. These studies examine if there are positive welfare gains to be made from improvements in recreational sites and thus help decision makers identify projects with positive net benefits. This in turn helps prioritize restoration projects under limited budgets.

A number of random utility maximization (RUM) models (Yen et. al, 1994; Herriges et al, 1997), count data models (Egan et al, 2006; Herriges et al, 2008) and Kuhn Tucker models (Phaneuf et al, 2000; Von Haefen et al, 2004) have been used to estimate the demand for outdoor recreation. Each of these models usually focuses on a different subset of characteristics of the recreation demand data. The Kuhn Tucker models are particularly suitable for studying site choice behavior when the consumer has the option to choose among several sites. In such cases, it is typical to observe a large number of consumers that do not visit any site at all. Those that do make a positive number of trips often make multiple visits to a small subset of the sites available. As a result, the data on the number of trips to recreational sites consists of a large number of zeros with the rest being positive numbers. The present paper examines the question of temporal stability of parameter estimates in a Kuhn-Tucker model using multiple years of data.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the precursors of the Kuhn-Tucker models and discusses the place and the importance of this approach in the context of recreation demand models. Section 3 sets up the model and describes the likelihood functions for different assumptions about the distributions of the error term. Section 4 presents a brief overview of the data. Section 5 presents the estimation results and the final section concludes.

Kuhn-Tucker Models – Background and Overview

The Kuhn-Tucker model is the latest in line of approaches that model the downward censored nature of recreation demand data, i.e the fact that trip demand cannot be negative. It is a development from the Amemiya-Tobin approach (Amemiya, 1974; Tobin, 1958) and the linked (participation and site choice) model (Bockstael et al., 1987). The Kuhn-Tucker formulations are improvements on the previous approaches in the sense that the Kuhn-Tucker model is utility consistent while the earlier approaches are not. Utility consistency implies that the decision to undertake a trip and the site choice decision are both derived from the same underlying utility function. However, till recently that these models were not extensively used because of their analytical intractability. Recent advances in computational resources have made it much easier to estimate these models leading to a number of studies in this area in recent years.

Phaneuf et al (2000) used the Kuhn Tucker model to estimate demand for fishing in the Wisconsin Great Lakes region and offered a method for estimating the expected welfare effects proxied by compensating variation associated with hypothetical policy changes in the Great Lakes region. The welfare estimation process involved computing the demands at every possible corner of the budget-constrained choice space and choosing the one that maximized utility. With the choice set consisting of four different sites, for each individual there were 16 different corners at which demands had to be estimated. This was computationally tractable. However, this method becomes unwieldy for large choice sets which are typical for many kinds of recreational choices. Von Haefen et al (2004) offered a method for estimating changes in welfare systems for large demand systems. All of these studies worked with data for a single year so that not much is known about the stability of the parameter estimates from the Kuhn-Tucker model over time. This is the gap that the present paper seeks to fill.

Model

For the model we assume that there are T periods and a total of M sites each having K characteristics. The following symbols are used.

x_{ijt} : Number of trips taken by the ith individual to site j in period t

$Q_t = [q_{1t}, q_{2t}, \ldots q_{Mt}]$: where q_{jt} is a K by 1 vector of quality variables associated with site j

S_{it} : The set of demographic characteristics for the ith household in period t

θ : Vector of nonuse values that the household assigns to each site

η_i : MT by 1 matrix of error terms for the ith individual. The term in the matrix indexed by ijt indicates the error associated with the jth site in period t,

z_{it} : A composite of all other goods (the numeraire and a necessary good)

γ and δ : Parameters of the model.

The utility function for a household i in period t is given by

$$U_{it} = \sum_{j=1}^{M} exp\left(\delta' S_{it} + \eta_{ijt}\right) ln\left[exp(\gamma' q_{it}) x_{ijt} + \theta\right] + \ln\,(z_{it}),$$

where η_{ijt} represent heterogeneity in preferences as the random components are assumed known to the individual but unobserved by the researcher. Associating the error term with preferences rather than with the demand makes this model consistent with McFadden's random utility maximization framework. This specification of the utility function is additively separable. Further, it assumes that every good (site in this case) is a normal good and all goods are Hicksian substitutes. The utility function also assumes weak complementarity meaning that quality attributes of a site do not affect the total utility of the individual if the site is not visited (Maler). In other words, the individual cares about the quality attributes of only those sites that s/he visits. The budget constraint for the individual in period t is

$$y_{it} = \sum_{j=1}^{M} P_{jt}\, x_{ijt} + z_{it};\ x_{ijt} \geq 0;\ \ z_{it} > 0\ \forall i, j, t.$$

The decision variables for individual i are x_{ijt} and z_{it}. The number of trips to any lake must be non-negative while the expenditure on the numeraire must be strictly positive. The Kuhn-Tucker first order condition for the ith individual, in period t with respect to the jth site is given by

$$\frac{\partial U_{it}}{\partial x_{it}} = exp\left[\delta' S_{it} + \eta_{ijt}\right] \frac{\exp(\gamma' q_{jt})}{\exp(\gamma' q_{jt}) x_{ijt} + \theta} \leq \frac{P_{jt}}{z_{it}}\ with\ complimentarity\ slackness\ .$$

This in turn implies that

$$\eta_{ijt} \leq ln\left[\frac{P_{jt}}{z_{it}}\right] + ln\left[x_{ijt} + \frac{\theta}{\exp(\gamma' q_{jt})}\right] - \delta' S_{it} = g\left(x_{it}, P_t, z_{it}, S_{it}, q_{jt}; \beta\right)$$

where β is the vector of parameters to be estimated. For each individual *i* in each period *t*, there are M such equations - one for each site. These equations together with the assumed distribution of the error term define a likelihood function which can then be used to estimate the parameters and welfare changes from hypothetical changes in site quality attributes.

Error Term

The error term for each individual and for each site are assumed to be correlated across time. The error term for the ith individual and jth site in period t consists of two components – one that is unique to the individual, the site and the time period and another which is unique to the individual and the site but remains constant over time. Errors for an individual are uncorrelated across sites and errors are uncorrelated across individuals. Formally, this can be written as:

$$\eta_{ijt} = u_{ij} + \varepsilon_{ijt}$$

$$E(u_{ij}|x) = e(\varepsilon_{ijt}|x) = 0$$

$$E(u_{ij}^2|x) = \sigma_u^2$$

$$E(\varepsilon_{ij}^2|x) = \sigma_\varepsilon^2$$

$$E(\eta_{ijt}^2|x) = \sigma_u^2 + \sigma_\varepsilon^2$$

$$E(\eta_{ijt}\eta_{ijs}|x) = \sigma_u^2 \forall t \neq s$$

$$E(\eta_{ijt}\eta_{ikt}|x) = 0 \; \forall j \neq k$$

$$E(\eta_{ijt}\eta_{kjs}|x) = 0 \; \forall \; i \neq k$$

This allows us to define the variance-covariance matrix for the error term. For the present paper, the errors are assumed to be drawn from a type I extreme value distribution.

Likelihood Function

The first order condition of the Kuhn-Tucker model together with the complimentarity slackness conditions require that

$$\eta_{ijt} = g(x_{it}, P_t, z_{it}, S_{it}, q_{jt}; \beta) \; if \; x_{ijt} > 0$$

$$\eta_{ijt} \leq g(x_{it}, P_t, z_{it}, S_{it}, q_{jt}; \beta) \; if \; x_{ijt} = 0$$

Let us assume that the ith consumer makes no trips to the first *n* sites and a positive number of trips to the rest. Then the likelihood function for the individual in each period is given by

$$L_i|\beta = \int_{-\infty}^{g(x_{i1t}, P_t, z_{it}, S_{it}, q_{1t}; \beta)} \dots \int_{-\infty}^{g(x_{i1t}, P_t, z_{it}, S_{it}, q_{1t}; \beta)} f(\eta_{it})|J|d\,\eta_{int} \dots d\eta_{int}$$

where $|J|$ is the Jacobian transformation of the error term and $f(\eta_{it})$ is the assumed probability density function of the error.

Data

For four consecutive years from 2002 to 2005, a survey questionnaire about trips to lakes in Iowa was sent out to a sample of households within the state. Each year about 4000 completed surveys were returned. However, some households dropped out of the sample for various reasons while others were added into it. A total of about 2000 individuals answered the questionnaire in all of the four years. For these 2000 individuals we have a balanced panel dataset for four years. For the purposes of this study we are using two of the four years of data. The questionnaire sought information about the actual and planned trips by the individuals to lakes in Iowa, some demographic information like age, gender, education, income and number of children in the family. A total of 132 lakes were covered in the survey. Of these we have information about water and site quality attributes in each of the four years for 127 lakes. The travel cost for each individual to each lake was computed using the travel time for the individual to the lake together with the prevailing price of gasoline and a fraction of the wages that the individual would have earned during that travel time. Table 1 presents a summary of the data on trips to the lakes.

Table 1: Trips to Lakes in Iowa

	2002	**2003**	**2004**	**2005**
0	367	572	647	433
1-10	663	626	653	603
11-25	356	315	274	382
26-40	158	101	79	156
41-55	70	52	34	64
55+	122	70	49	98
Total	1736	1736	1736	1736

A sizeable fraction of the sample did not make any trips to the lakes at all. The largest fraction of the sample made between one and ten trips. One point that is not evident from Table 1 is that people who made many trips, did not visit all or even a majority of lakes. They made multiple visits to a small subset of lakes.

Table 2 presents certain demographic characteristics of the sample.

Table 2: Demographic Information

	2002		2003	
	Mean	Std. Dev	Mean	Std. Dev
Income	27068	5987	27345	5965
Age	53.7	14.9	54.4	14.9
Sex*	1.3	0.4	1.3	0.4

*: Male = 1; Female = 2

The average income for the sample was marginally over $27,000 while the mean age was around 54 years. The demographic characteristics of the sample look almost identical in the two years because the sample in the two years consists of the same individuals.

Table 3 presents statistics on select water quality variables of the 127 lakes in our dataset. These variables are Secchi depth, which is a measure of clarity, chlorophyll and phosphorus which are pollutants in the water. There was a substantial amount of variation in the average water quality of the lakes in the two years. There appears to have been some improvement in water quality in 2003. The mean Secchi depth increased as compared with 2002 while the mean chlorophyll and phosphorus loads decline substantially.

Table 3: Water Quality Variables						
	2002			2003		
	Secchi Depth (m)	Chlorophyll (ug/l)	Phosphorus (ug/l)	Secchi Depth (m)	Chlorophyll (ug/l)	Phosphorus (ug/l)
Mean	1.2	40.4	105.3	1.5	20.0	94.0
Std Dev	0.9	38.1	80.8	1.1	7.8	66.4
Max	5.7	182.9	452.6	8.1	37.6	383.8
Min	0.1	2.5	17.1	0.2	2.1	16.9

Results

The Kuhn-Tucker model was estimated for 2002 and 2003 assuming that the errors are distributed as a type 1 extreme value distribution. Table 4 presents the parameter estimates for 2002 and 2003. Most of the estimates appear to be fairly stable across the two years. The demographic variables have the expected signs and are highly significant in both periods. The magnitude of the constant term appears to be substantially smaller in 2003 as compared with the previous year.

Table 4: Parameter Estimates		
Estimates for 2002		
Parameter	**Estimates**	**Standard Error**
Constant[1]	-8.91**	0.08
Age	-0.21**	0.01
Gender[2]	-0.05*	0.02
Secchi Depth (m)	0.04*	0.02
Chlorophyll (ug/l)	-0.58	0.02
Total Phosphorus (ug/l)	0.57	2.00
Theta (Non-use value)	1.75**	0.05

Mean Log Likelihood	-0.1315	
Estimates for 2003		
Parameter	**Estimates**	**Standard Error**
Constant[1]	-7.46**	0.07
Age	-0.13**	0.01
Gender[2]	-0.01	0.01
Secchi Depth (m)	0.10**	0.01
Chlorophyll (ug/l)	0.60	8.40
Total Phosphorus (ug/l)	-0.61	8.40
Theta (Non-use value)	2.10**	0.07
Mean Log Likelihood	-0.1689	

1: Constant related to demographic attributes. 2: Male = 1; Female = 2. *: Significant at 95% confidence level. **: Significant at 99% confidence level.

The estimated coefficients of the water quality variables appear to be less stable. The coefficient of Secchi depth is positive and significant in both years though the estimates appear to differ over the two years. The estimates for the coefficients of total phosphorus and chlorophyll are different for the two years and are of the wrong sign (positive) in one of the two years. However, neither is significant in 2002 or 2003. The parameter of non-use value is of the right sign and highly significant. Overall, the estimates indicate a certain stability of estimates over the two consecutive years.

Conclusion

This study presented an estimation of the Kuhn-Tucker model with a large dataset for two consecutive years. The results indicate that for our specific dataset, the estimates are temporally stable. The significant difference in the average water quality of the lakes between 2002 and 2003 adds credibility to this result. The next step in the process is to estimate a panel data model with these two years to see if there are any gains in efficiency from the joint estimation.

References

Amemiya, T., “Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal,” Econometrica 42 (September 1974), 999-1012.

Bockstael, N.E., W.M. Hanemann, and C.L. Kling, “Estimating the Value of Water Quality Improvements in a Recreation Demand Framework,” *Water Resources Research* 23 (May 1987), 951-60.

Bockstael, N.E., W.M. Hanemann, and I.E. Strand, *Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models*, Vol. 2, Washington D.C.: U.S. Environmental Protection Agency, Office of Policy Analysis (1986).

Egan, Kevin J., Joseph A. Herriges, Catherine L. Kling, John A. Downing. “Recreation Demand Using Physical Measures of Water Quality” *American Journal of Agricultural Economics*, Vol. 91, 1, pp. 106-123, 2009

Herriges, J.A., and C.L. Kling. “Performance of Nested Logit Models when Welfare Estimation Is the Goal (The),” *American Journal of Agricultural Economics* , Vol. 79, 3, pp. 792-802, August, 1997

Herriges, J.A., C.L. Kling, and D.J. Phaneuf, “What’s the Use? Welfare Estimates from Revealed Preference Models when Weak Complementarity Doesn’t Hold,” *Journal of Environmental Economics and Management* 47 (2004): 55-70.

Joseph A., Daniel J. Phaneuf, Justin Tobias. “Estimating Demand Systems when Outcomes are Correlated Count” *Journal of Econometrics* , Vol. 147, 2, pp. 282-298, 2008

Maler, K., *Environmental Economics*, Baltimore: Johns Hopkins, 1974.

McFadden, D., "Econometric Models of Probabilistic Choice," in Charles F. Manski and Daniel L. McFadden (eds.), *Structural Analysis of Discrete Choice Data* (Cambridge, MA: MIT Press, 1981).

Morey, E.R., “The Choice of Ski Areas: Estimation of a Generalized CES Preference Ordering with Characteristics, Quadratic Expenditure Functions, and Non-Additivity,” *Review of Economics and Statistics* 66 (1984), 221-36.

Morey, E.R., W.D. Shaw, and R.D. Rowe, “A Discrete Choice Model of recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data Are Not Available,” *Journal of Environmental Economics and Management* 20 (1991), 181-201.

Morey, E.R., D. Waldman, D. Assane, and D. Shaw, “Specification and Estimation of a

Generalized Corner Solution Model of Consumer Demand: An Amemiya-Tobin Approach," working paper, May 1990.

Morey, E.R., D. Waldman, D. Assane, and D. Shaw, "Searching for a Model of Multiple-Site Recreation Demand that Admits Interior and Boundary Solutions," *American Journal of Agricultural Economics* 77 (1995), 129-140.

Tobin, J., "Estimation of Relationships for Limited Dependent Variables, *Econometrica* 26 (1958), 24-36.

von Haefen, R., D. Phaneuf, and G. Parsons. "Estimation and welfare calculation with large demand systems," *Journal of Business and Economic Statistics* 22 (2004): 194-205.

Yen, S. T., and W. L. Adamowicz, "Participation, Trip Frequency, and Site Choice: A Multinomial Poisson Hurdle Model of Recreation Demand," *Canadian Journal of Agricultural Economics* 42 (1994), 65-76.