DRAFT: May 2001

Decision Sequence in the Nested-Logit Model of Recreation Choice: An Application to Oregon Marine Sport Fishing

Sabina L. Shaikh
(contact author)
Assistant Professor
Faculty of Agricultural Sciences
University of British Columbia
248-2357 Main Mall
Vancouver, BC V6T 1Z4, Canada
Phone/Fax: (604) 822-2144
Email: sshaikh@interchange.ubc.ca

and

Douglas M. Larson
Associate Professor
Department of Agricultural and Resource Economics
University of California-Davis

Selected Paper for the 2001 American Agricultural Economics Association Meetings
August 5-8, Chicago, IL

Abstract:

The varying researcher assumptions about model specification in nested-logit random utility models can have significant effects on model estimates and corresponding welfare measures. Using a three-level nested-logit model of Oregon marine sport fishing, we estimate and compare several models with varying assumptions of choice sequence and income effects.

I. Introduction

There has been significant treatment of the welfare effects for changes in prices and quality related to recreation activities. Welfare measurement has been shown to be sensitive to various assumptions used in recreation demand modeling. For example, the order in which individuals make decisions is an important element of consumer decision modeling, which may significantly affect model results (Kling and Thomson (1996), Shaw and Ozog (1999), Hauber and Parsons (2000).) Welfare estimates may also be sensitive to assumptions about income, particularly in the case of random utility models. This research seeks to examine Oregon marine sport fishing behavior and sensitivity of welfare estimates to various specifications of nesting order and income in the nested-logit random utility model (RUM).

Random utility models have proven to be a useful and effective tool for modeling recreation choices made from a set of available options. The nested-logit is the most commonly-used due in part to its ability to model several decision levels, which may or may not be independent of each other. In the case of a fishing trip, the overall decision may involve three choices, where to fish, how to fish and what to fish for. The order in which an individual makes these choices, may have nontrivial implications for the accuracy of welfare measures for policy-induced changes in quality or accessibility. While researchers usually make an arbitrary assumption about the order in which individuals make the decisions that comprise an overall experience, this assumption is not generally tested and it is possible that a different sequence could yield dramatically different results (Hauber and Parsons (2000)).

Also, since RUM's generally rely on linear interpretations of income, welfare estimates are often subject to the assumption of zero income effects. Furthermore, because recreation is a time-intensive activity, additional assumptions are required to incorporate time prices and budgets into models. Clearly, welfare estimation is subject to several researcher assumptions, and isolating the effects of each is a difficult task. Using recreational fishing data from the National Marine Fishing Service (NMFS) for Oregon marine sport anglers, we construct a three-level random utility nested-logit model of site, mode and target species choice. We estimate several model specifications, which vary the assumptions of decision sequence, income effects and time values. Welfare measures will be calculated for each model and compared.

The paper is organized as follows. Section II provides a brief overview of the random utility model framework. The data are described in Section III, followed by the empirical model specification and econometric model in Section IV. Section V describes the undergoing estimation and expected results.

II. The Random Utility Model

Because recreation is a time-intensive good, the model is based on a two-constraint utility maximization model. A random utility model of fishing choice with is used as the empirical specification. We begin with the common linear-in-the-parameters specification and extend it to a model, which allows for nonzero income effects. RUM's with nonlinear income effects have implemented used in recent literature to relax the restrictive, although convenient assumption of a constant marginal utility of income across alternatives (e.g. Herriges and Kling (1999), Shaw and Ozog (1999). However, the appropriate theoretical specification of the time parameters becomes increasingly important in nonlinear RUMs but is often overlooked through the inclusion of time prices without a leisure time budgets. This leads inconsistency of the empirical model with the underlying two-constraint utility maximization problem (Larson and Shaikh (2000).

While there are several variants of the RUM, the standard model for a single choice takes the form:

(1) $$U_j = V_j + \varepsilon_j \qquad j = 1 \text{ to } J$$

where j represents the alternative chosen out of J possible alternatives. Utility is comprised of two elements, a systematic component denoted by V_j and a random component ε_j to capture certain factors unobservable to the researcher. The vector of J random terms is given as

(2) $$\varepsilon_j = \{\varepsilon_1, \varepsilon_2, \ldots\ldots., \varepsilon_J\} \text{ where}$$

$f(\varepsilon_j)$ is the PDF and $F(\varepsilon_j)$ is the CDF.

A comparison is made between alternatives and an individual will choose the one which yields the highest level of utility. The probability of choosing alternative m, for example, is given by

(3) $$\begin{aligned} \Pr ob(m) &= \Pr ob\left[V_m + \varepsilon_m > V_j + \varepsilon_j \quad \forall m \neq j\right] \\ &= \Pr ob\left[\varepsilon_j < V_m - V_j + \varepsilon_m \quad \forall m \neq j\right] \end{aligned}$$

The log-likelihood function for estimation is then given as:

(4) $$L = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} \ln\left[\Pr ob(ij)\right]$$

The model can easily be extended for several choices and formulated as a nested-logit model by using a GEV distribution of errors. It is then estimated sequentially or more commonly using full-information maximum likelihood. The empirical model and estimation techniques will be discussed following the data description. A good description of the nested-logit models and their estimation is given in Morey (2000).

III. Oregon Sport Fishing Data

The data is from the 1998 Pacific States Marine Recreational Fishery Statistical Survey (MRFSS) from the NMFS. The data is from an on-site intercept survey over six fishing waves (one calendar year), combined with a telephone add-on survey. Only those for whom an add-on survey was conducted are used for this paper.

For the state of Oregon, there are seven possible sites located in Clatsop, Coos, Curry, Douglas, Lane, Lincoln and Tillamook counties. There are four possible modes for fishing–two shore modes (manmade and beach)–and two boat modes (charter and private), and six possible target species categories–unidentified, groundfish, sturgeon, salmon, perch and all others. Descriptive statistics for personal characteristics of anglers given in Table 1 are money income, leisure time budget (total available time–work time), hours worked per week, age and education level. The response rate for income was low–79%, which is a not unusual in this type of survey (Whitehead and Haab (2000)). Therefore, following Whitehead and Haab's approach for the MRFSS southeast data, income was imputed by running an OLS regression the available observations of log household income, with the log of census zip code income on the right hand side, in addition to gender, age, boat ownership and job status. Missing observations for income were then imputed using the regression results.

Table 2 gives information pertaining to fishing experience including the number of day trips taken in the last two months, number of overnight trips in the past two months and years of saltwater fishing experience. Tables 3-6 provide descriptive statistics and information about the fishing trip on which the individual was surveyed. Tables 3-5 provide the basis for the nested-logit random utility model and indicates the percentage of the sample that chose specific sites, modes and species. Table 6 gives additional information about the fishing trip, in particular, the costs associated with a trip. Travel miles and distances were determined using the PC-Miler computer package to calculated the distance between the individual's home zip code and the zip code of the fishing site. Other expenses for the trip are the gear and bait expense, a daily license fee and any boat and parking expenses.

Five-year historical catch rates are used as a proxy for fishing quality. Using the catch rates by species, by mode, by site for a particular wave, the five-year average catch per unit of effort was calculated to represent quality. Expected catch rates are another possibility for the quality representation, but are not used in order to avoid introducing additional potential sources of econometric error into the already complicated model.

Each observed result for an individual is assumed to be a final decision based on a series of choices made from an individual-specific choice set. There are 168 possible alternatives but only 96 available options due to not all species being available by mode or site and not all modes being available at each site. The six possible nesting options to be explored here are

Possible Nesting Structures (168 Total Possibilities: 96 Available Choices)

Site Mode Species	Site Species Mode	Species Mode Site	Species Site Mode	Mode Site Species	Mode Species Site

IV. Empirical and Econometric Model

A three-level nested-logit model is specified based on a two-constraint utility maximization problem. The random utility model for the choice in the order of site, mode and speices is given as

$$U_{jkl} = V_{jkl} + \varepsilon_{jkl} \qquad \begin{array}{l} j = 1 \text{ to } 7 \textit{ sites} \\ k = 1 \text{ to } 4 \textit{ modes} \\ l = 1 \text{ to } 6 \textit{ species} \end{array} \tag{5}$$

Using Morey's (2000) notation, for the three-level nested-logit model, without loss of generality, the probability of choosing site 2, mode 3 and species 4, is

$$\text{prob}(234) = \frac{\left(\exp(sV_{234}) \left[\sum_{k=1}^{4} \left[\sum_{l=1}^{6} \exp(sV_{2kl}) \right]^{t/s} \right]^{(1/t)-1} \right) \cdot \left[\sum_{j=1}^{6} \exp(sV_{23j}) \right]^{((t/s)-1)}}{\sum_{j=1}^{7} \left[\sum_{k=1}^{4} \left[\sum_{l=1}^{6} \exp(sV_{jkl}) \right] \right]} \tag{6}$$

where s and t represent the degrees of substitution across alternatives in the third-level and second-level nests, respectively. In the case that s = t, the model reduces to a non-nested multinomial choice model. Note that this is a simplified example, which assumes that the s and t are constant across different alternatives. In practice, s and t will vary with the number of alternatives available at each nesting level. Hauber and Parsons (2000) find that these parameters are quite important in determining the significance of decision sequence on resulting welfare estimates.

The choice probability in (6) is representative of a particular decision sequence, as demonstrated by the denominator, which is the inclusive value. In order to test different decision sequences, the nests can be switched around by changing the order of the inclusive value and adjusting the inclusive parameters to indicate the number of alternatives in each nest level.

In addition to decision sequence, the specification of income within the RUM can potentially have significant effects on welfare measurement. While Hauber and Parsons and Kling and Thomson test decision sequence, they use models which zero income effects. Shaw and Ozog use a nonlinear income specification but are unable to estimate model parameters under different decision sequences. While its not entirely clear how income is defined, it appears that they include time prices without corresponding time budgets in a RUM where income does not drop out of the choice probabilities (i.e. nonlinear income effects), which can lead theoretical model inconsistencies (Larson and Shaikh (2000)). Morey, *et al*, do use both full prices and full budgets in their repeated nested-logit model, however, all individuals are assumed to have the same time budget, which is unlikely.

We use both a linear and nonlinear specification, both of which use a construction of *full* prices and *full* budgets, converted by the marginal value of time ρ.[1] Assuming

[1]The appropriate specification of ρ has undergone significant treatment in the literature, the most general and widely used is an endogenous or exogeneous fraction of the wage rate.

again that the choice order is site, mode and species, the following are the specifications for indirect utility.

Linear Indirect Utility:

(7) $$V_{jkl} = \alpha(M^F - p^F_{jkl}) + \delta(cr_{jkl}) + \hat{\mathbf{a}}\mathbf{s}$$,where

cr_{jkl}: catch rate for species l by mode k at site j

M^F: full annual income comprised of money income plus the leisure time budget, converted to dollars by the marginal value of time ρ.

p^F_{jkl}: full price of a trip to site j, by mode k for species l, where price is comprised of travel costs plus additional expenses plus travel time times the marginal value of time ρ.

s: vector of individual characteristics (age, education, saltwater fishing experience, etc.)

Relaxing the assumption of zero income effects complicates welfare estimation slightly, but allows for nonlinear income effects. A modification of (7) gives

Nonlinear Indirect Utility

(8) $$V_{jkl} = \alpha(M^F - p^F_{jkl}) + \alpha_2(M^F - p^F_{jkl})^2 + \delta(cr_{jkl}) + \hat{\mathbf{a}}\mathbf{s}$$

This is a simple modification designed to test only the effect of nonlinear income. A more flexible model, such as a translog, would allow for nonlinear effects of catch rates and individual characteristics as well (Herriges and Kling (1999)).

Using the indirect utility specifications in (7) and (8) in the choice probabilities in (6), the log-likelihood function for estimation is

(9) $$L = \sum_{i=1}^{N} \sum_{j=1}^{7} \sum_{k=1}^{4} \sum_{l=1}^{6} y_{ijkl} \ln[\Pr ob(ijkl)]$$, where

i is the index for individuals. In this case there are N=3195 individuals and 96 total choices alternatives. The model is estimated by full information maximum likelihood estimation using the Maximum Likelihood Module in Gauss 3.5

Welfare measurement for the linear model is straightforward. For the nonlinear model, analytical solutions are not possible, however, several options are available for welfare calculation. Morey, *et al* and Shaw and Ozog employ numerical approximations for a representative consumer. McFadden (1999) develops a Monte Carlo Markov Chain simulation method, which does not rely on numerical methods for welfare measurement. In a recent and comprehensive treatment, Herriges and Kling modify this algorithm to identify bounds for welfare measures when there are nonlinear income effects.

Estimation is underway and preliminary are promising, but not yet complete. In addition into uncovering valuable information and policy implications for Oregon marine sport fishing, the objective of this research is to identify the various factors or combination of factors that lead to variations in random utility welfare estimation. This paper will be updated shortly to include estimation results and inference. We greatly appreciated the reader’s patience and welcome any comments or suggestions.

Tables

Table 1: Personal Characteristics of Sample

Variable	Mean	Std Dev	Variance	Min	Max
Imputed Income ($)	42,900	63,675.46	4054.00e+6	1,349	1,070,900
Hours worked/week	44.27	10.85	117.63	2.00	80.00
Leisure Time Budget	6617.55	665.020	442,246.34	4760	8860
Age (years)	47	15.25	232.57	8	92
Highest Education Attained		**Percent of sample**			
<12 years		13 %			
high school grad		36 %			
some college or trade school		21 %			
2 year college degree		13 %			
4 year college degree		12 %			
postgraduate degree		6 %			

Table 2: Characteristics pertaining to marine fishing

Variable	Mean	Std. Dev	Variance	Min	Max
Day trips in past 2 mo (#)	8	9.12	82.98	0	60
Overnight trips in past 2 mo (#)	1.8	0.49	0.25	1	9
Years of saltwater fishing experience	18.99	15.55	241.69	1	75

Table 3: Site Chosen by Individuals

County (Site)	Percentage of Sample
1. Clatsop	4 %
2. Coos	13 %
3. Curry	32 %
4. Douglas	10 %
5. Lane	1 %
6. Lincoln	23 %
7. Tillamook	17 %

Table 4: Mode Chosen by Individuals

Mode Type	Percentage of Sample
1. Manmade Dock or Pier	11 %
2. Beach/Bank	15 %
3. Party/Charter Boat	20 %
4. Private/Rental Boat	54 %

Table 5: Species Targeted by Individuals

Species Category*	Percentage of Sample
1. Unidentified	9 %
2. Bottom	39 %
3. Sturgeon	10 %
4. Salmon	31 %
5. Perch	7 %
6. Other Targeted Species	4 %

*Note: species were aggregated into general categories. Bottom fish comprises all groundfish including rockfish and lingcod. "Unidentified" is a given response for those who did not target a particular species. "Other" refers to those respondents who identified a target species other than the ones listed in the table. These were aggregated into one "other" category due to the small number of individuals targeting individual species.

Table 6: Cost per trip

Variable	Mean	Std. Dev	Variance	Min	Max
Distance Traveled to Site (miles)	47.53	63.40	4019.87	1	802.5
Time Traveled to Site (hours)	1.09	1.22	1.48	0.1	14.6
Gear and Bait Expense ($)	4.25	12.17	148.11	0	300
Daily License Expense ($)	1.36	3.53	12.43	0	44
Boat and Parking Expense ($)	16.89	41.11	1690.38	0	400

References

Hauber, A. B and G. R. Parsons. "The Effect of Nesting Structure Specification on Welfare Estimation in a Random Utility Model of Recreation Demand: An Application to the Demand for Recreational Fishing." *American Journal of Agricultural Economics,* 82 (August 2000): 501-514.

Herriges, J. A. and C. L. Kling. "Nonlinear Income Effects in Random Utility Models." *Review of Economics and Statistics*. 81 (1999): 62-72.

Kling, C. L. and C. J. Thompson. "The Implications of Model Specification for Welfare Estimation in Nested Logit Models." *American Journal of Agricultural Economics*, 78 (February 1996): 103-114.

Larson, D. M. and S. L. Shaikh. "Empirical Specification Requirements for Two-Constraint Models of Recreation Choice." *American Journal of Agricultural Economics*. 85 (May 2001): 428-440.

McFadden, D. M. "Computing Willingness-to-Pay in Random Utility Models," in *Trade, Theory and Econometrics: Essays in Honor of John S. Chipman*. Routledge: London, 1999.

Morey, E. R. "TWO RUMs uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice in *Valuing Recreation and the Environment*, J. A. Herriges and C. L. Kling, eds. Northampton, MA: Edward Elgar, 2000.

Morey, E. R., R. D. Rowe and M. Watson. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing." *American Journal of Agricultural Economics*. 75 (August 1993): 578-592.

Shaw, W. D. and M. T. Ozog. "Modeling Overnight Recreation Trip Choice: Application of a Repeated Nested Multinomial Logit Model." *Environmental and Natural Resource Economics*. 13 (1999): 397-414.

Whitehead, J. C. and T. C. and Haab. "Southeast Marine Recreational Fishery Statistical Survey: Distance and Catch Based Choice Sets." *Marine Resources Economics*. 14 (2000): 283-298.