MEMORANDUM | September 1, 2015

TO Craig O'Connor, NOAA
FROM Joseph Herriges
SUBJECT D1 - Model Structure

The purpose of this memo is to describe the modeling structure used to characterize recreation demand. This model is then used in estimating the value per lost user day and constructing the damage estimate (See Technical Memorandum A – Overview of Recreation Assessment). The overall framework employed is the Repeated Logit (RL) model, developed by Morey, Rowe, and Watson [6] and based on the Random Utility Maximization (RUM) construct (Marschak [1], McFadden, [2, 3, 4], Thurstone [8]). We begin with a review of the generic RUM framework.

THE GENERIC RUM FRAMEWORK

RUM models assume that individuals, facing a well-defined choice set, select the alternative yielding the highest level of utility. Thus, if U_{ik} denotes the conditional utility received by individual i in choosing alternative k ($k = 1, \ldots, J$), then the individual chooses alternative j (denoted by $y_{ij} = 1$) if $U_{ij} > U_{ik}\ \forall k \neq j$.

$$y_{ij} = \begin{cases} 1 & U_{ij} > U_{ik}\ \forall k \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The conditional utilities themselves can depend upon characteristics of both the individual and the available alternatives.

The analyst modeling observed choices in a given setting will not observe all of the factors influencing the individual's decisions. Instead, they characterize the conditional utilities as a function $V_{ij} = V(\boldsymbol{X}_{ij}; \beta)$ of observable individual/alternative specific attributes ($\boldsymbol{X}_{ij}$), where β denotes a vector of parameters to be estimated, and a residual term ϵ_{ij}, implicitly defined as $\epsilon_{ij} = U_{ij} - V_{ij}$, which captures unobserved factors influencing the utility individual i derives from choosing alternative j. Thus $U_{ij} = V_{ij} + \epsilon_{ij}$. Given assumptions regarding the distribution of the vector $\epsilon_{i\bullet} \equiv (\epsilon_{i1}, \ldots, \epsilon_{iJ})$, the analyst can then identify the probability that a specific choice will be made. In general,

$$\begin{aligned} P_{ij} &= Pr(y_{ij} = 1 | \boldsymbol{X}_{i\bullet}) \\ &= Pr\left(U_{ij} > U_{ik}\ \forall k \neq j\right) \\ &= Pr\left(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}\ \forall k \neq j\right), \end{aligned} \tag{2}$$

where $\boldsymbol{X}_{i\bullet} = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iJ})$. These probabilities can, in turn, be used to specify the appropriate log-likelihood function used in maximum likelihood estimation of the model parameters.

THE REPEATED LOGIT MODEL

The Repeated Logit Model was developed by Morey, Rowe and Watson [6] in the context of recreation demand. There are two key modifications to the basic RUM model. First, a "stay-at-home" option is added to the choice set, so that j now runs from 0 to J, with $j = 0$ denoting the "stay-at-home" option and $j = 1,\ldots,J$ denoting the available sites in the choice set. Second, instead of making a single choice, individuals are modeled as choosing from among the $J+1$ alternatives over a series of T_i choice occasions. The number of choice occasions varies over individuals in our analysis to the extent that they have different reporting periods for trips.

The utility that individual i receives from choosing alternative j on choice occasion t typically takes the form:

$$U_{ijt} = V_{ij} + \epsilon_{ijt} \;\; j = 0,\ldots,J; t = 1,\ldots,T_i; i = 1,\ldots,N. \tag{3}$$

Assuming that the vector $\epsilon_{i\bullet t} \equiv (\epsilon_{i1t},\ldots,\epsilon_{iJt})$ is identically distributed across choice occasions, the probability of observing alternative j being chosen by individual i on choice occasion t (denoted by $y_{ijt} = 1$ if the alternative is chosen; $= 0$ otherwise) becomes:

$$\begin{aligned} P_{ijt} &= Pr(y_{ijt} = 1|\mathbf{X}_{i\bullet}) \\ &= Pr\left(U_{ijt} > U_{ikt} \;\; \forall k \neq j\right) \\ &= Pr\left(\epsilon_{ikt} - \epsilon_{ijt} < V_{ij} - V_{ik} \;\; \forall k \neq j\right) \\ &= P_{ij} \;\; \forall t = 1,\ldots,T_i, \end{aligned} \tag{4}$$

where the last equality follows from the fact that $\epsilon_{i\bullet t}$ is identically distributed across choice occasions and the V_{ij} are not a function of the choice occasion. Assuming that the error vector is also independently distributed across choice occasions, individual i's contribution to the log-likelihood function becomes:

$$\begin{aligned} \mathcal{L}_i &= \sum_{j=0}^{J}\sum_{t=1}^{T_i} y_{ijt} ln(P_{ijt}) \\ &= \sum_{j=0}^{J}\sum_{t=1}^{T_i} y_{ijt} ln(P_{ij}) \\ &= \sum_{j=0}^{J} ln(P_{ij}) \sum_{t=1}^{T_i} y_{ijt} \\ &= \sum_{j=0}^{J} n_{ij} ln(P_{ij}), \end{aligned} \tag{5}$$

where

$$n_{ij} = \sum_{t=1}^{T_i} y_{ijt} \tag{6}$$

denotes the number of times individual i chooses alternative j over their T_i choice occasions. Completing the model requires specifications for the V_{ij}'s and distributional assumptions for the error vectors (i.e., the $\epsilon_{i\bullet t}$'s).

SPECIFYING THE V_{ij}'s

The V_{ij}'s are assumed to have the following structure:

$$V_{ij} = \begin{cases} \delta Z_i & j = 0 \\ \alpha_j + \beta C_{ij} & j = 1, \ldots, J \end{cases} \tag{7}$$

where the Z_i denote individual-specific characteristics impacting the individual's propensity to choose the "stay-at-home" option ($j = 0$) and C_{ij} denotes the roundtrip travel cost for individual i in choosing to visit alternative j on a given choice occasion. The α_j parameters, commonly referred to as *alternative-specific constants (ASC's)* in the literature (e.g., Murdock [7]), capture the site-specific attributes of alternative j, while β captures the impact that the cost of visiting a site has on the propensity to visit a site. It is commonly referred to as the negative of the marginal utility of income.

THE ERROR DISTRIBUTION

The final step in completing the model specification is to choose the error distribution for the vector $\epsilon_{i\bullet j}$.

THE TWO-LEVEL NESTED LOGIT (NL2)

The most commonly used structure in the recreation demand literature is to assume that the $\epsilon_{i\bullet t}$ are drawn from a GEV distribution implying a two-level nesting structure that groups the trip alternatives (i.e., j = 1, . . . , J) into a nest and the stay-at-home option into a singleton nest. The implied choice probabilities then take the form:

$$P_{ij} = \begin{cases} \dfrac{exp(V_{i0})}{exp(V_{i0}) + \left[\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)\right]^{\theta}} & j = 0 \\ \dfrac{exp\left(\frac{V_{ij}}{\theta}\right)}{\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)} \cdot \dfrac{\left[\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)\right]^{\theta}}{exp(V_{i0}) + \left[\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)\right]^{\theta}} & j = 1, \ldots, J \end{cases} \tag{8}$$

$$= \begin{cases} 1 - P_{i,Trip} & j = 0 \\ P_{ij|Trip} \cdot P_{i,Trip} & j = 1, \ldots, J \end{cases}$$

where θ is often referred to in the literature as the dissimilarity coefficient,

$$P_{i,Trip} = \frac{\left[\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)\right]^{\theta}}{exp(V_{i0}) + \left[\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)\right]^{\theta}} \tag{9}$$

denotes the probability that individual i chooses to take a trip on a given choice occasion and

$$P_{ij|Trip} = \frac{exp\left(\frac{V_{ij}}{\theta}\right)}{\sum_{k=1}^{J} exp\left(\frac{V_{ik}}{\theta}\right)}$$

denotes the probability that they choose to visit site j ($j = 1,\ldots,J$) conditional on having chosen to take a trip.

THREE-LEVEL NEST (NL3)

The Three-Level Nested Logit (NL3) generalizes the NL2 by further dividing the *Trip* nesting structure into sub-groups. Specifically, suppose that the J trip sites are assigned to one of G sub-groups. The set of sites contained in sub-group g ($g = 1,\ldots, G$) is denoted by N_g, with $g(j)$ denoting the sub-group that site j belongs to. Thus, if site 2 belongs to sub-group 3, then $g(2) = 3$ and $2 \in N_3$. The corresponding choice probabilities become:

$$P_{ij} = \frac{exp\left(\frac{V_{ij}}{\theta\tau_{g(j)}}\right)}{\sum_{k\in N_{g(j)}} exp\left(\frac{V_{ik}}{\theta\tau_{g(j)}}\right)} \cdot \frac{\left[\sum_{k\in N_{g(j)}} exp\left(\frac{V_{ik}}{\theta\tau_{g(j)}}\right)\right]^{\tau_{g(j)}}}{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}} \cdot \frac{\left\{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}\right\}^{\theta}}{exp(V_{i0}) + \left\{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}\right\}^{\theta}}$$

$$= P_{ij|g(j)} \cdot P_{i,g(j)|Trip} \cdot P_{i,Trip}$$

for $j = 1,\ldots,J$ and

$$P_{i0} = \frac{exp(V_{i0})}{exp(V_{i0}) + \left\{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}\right\}^{\theta}} = 1 - P_{i,Trip},$$

where

$$P_{ij|g(j)} = \frac{exp\left(\frac{V_{ij}}{\theta\tau_{g(j)}}\right)}{\sum_{k\in N_{g(j)}} exp\left(\frac{V_{ik}}{\theta\tau_{g(j)}}\right)} \quad (11)$$

denotes the probability that site j is chosen, given that the sub-nest that site j is in (i.e., $g(j)$) has been chosen,

$$P_{i,g(j)|Trip} = \frac{\left[\sum_{k\in N_{g(j)}} exp\left(\frac{V_{ik}}{\theta\tau_{g(j)}}\right)\right]^{\tau_{g(j)}}}{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}} \quad (12)$$

denotes the probability that sub-nest $g(j)$ has been chosen, given that the individual has decided to take a trip, and

$$P_{i,Trip} = \frac{\left\{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}\right\}^{\theta}}{exp(V_{i0}) + \left\{\sum_{h=1}^{G}\left[\sum_{l\in N_h} exp\left(\frac{V_{il}}{\theta\tau_h}\right)\right]^{\tau_h}\right\}^{\theta}} \qquad (13)$$

denotes the probability that the individual decides to take a trip on a given choice occasion. A variation on the above model commonly used in the literature is to assume that $\tau_g = \tau \ \forall g = 1, \ldots, G$; i.e., that the sub-nest dissimilarity coefficients are the same across the sub-nests.

REFERENCES

[1] Marschak, Jacob. 1960 Binary Choice Constraint on Random Utility Indications. In Stanford Symposium on Mathematical Methods in the Social Sciences, 312319, edited by K. Arrow. Stanford: Stanford University Press.

[2] McFadden, Daniel. 1973. Conditional Logit Analysis of Qualitative Choice Behav- ior. In Frontiers of Econometrics, edited by Paul Zarembka. New York: Academic Press.

[3] McFadden, Daniel. 1978. Modeling the Choice of Residential Location. In Spa- tial Interaction and Planning Models, 7596, edited by A. Karlqvist. Amsterdam: North Holland.

[4] McFadden, Daniel. 1981. Econometric Models of Probabilistic Choice Mod- els. In Structural Analysis of Discrete Choice Data with Applications, edited by Charles Manski and Daniel McFadden. Cambridge: MIT Press.

[5] Morey, Edward R. 1999. Two RUMs Uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice. In Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice, edited by Joseph A. Herriges and Catherine L. Kling Cheltenham: Ed- ward Elgar.

[6] Morey, Edward R., Robert D. Rowe, and Michael Watson. 1993. A Repeated Nested- Logit Model of Atlantic Salmon Fishing. American Journal of Agricul- tural Economics 75(3): 578592.

[7] Murdock, Jennifer. 2006. Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand. Journal of Environmental Economics and Management, 51(1): 125.

[8] Thurstone, Louis. 1927. A Law of Comparative Judgement. Psychological Review. 34: 273 286.