WESTERN REGIONAL RESEARCH PUBLICATION

W-133
BENEFITS AND COSTS OF RESOURCES POLICIES AFFECTING PUBLIC AND PRIVATE LAND

12TH INTERIM REPORT
JUNE 1999

Compiled by
W. Douglass Shaw

Department of Applied Economics and Statistics
Mail Stop 204
University of Nevada
Reno, Nevada 89557-0105

INTRODUCTION

This volume contains the proceedings of the 1999 W-133 Western Regional Project Technical Meeting on "Benefits and Costs of Resource Policies Affecting Public and Private Land." Some papers from W-133 members and friends who could not attend the meeting are also included. The meeting took place February 24th - 26th at the Starr Pass Lodge in Tucson, Arizona. Approximately 50 participants attended the 1999 meeting, are listed on the following page, and came from as far away as Oslo, Norway.

The W-133 regional research project was rechartered in October, 1997. The current project objectives encourage members to address problems associated with: 1.) Benefits and Costs of Agro-environmental Policies; 2.) Benefits Transfer for Groundwater Quality Programs; 3.) Valuing Ecosystem Managment of Forests and Watersheds; and 4.) Valuing Changes in Recreational Access.

Experiment station members at most national land-grant academic institutions constitute the official W-133 project participants. North Dakota State, North Carolina State, and the University of Kentucky proposed joining the group at this year's meeting. W-133's list of academic and other "Friends" has grown, and the Universities of New Mexico and Colorado were particularly well represented at the 1999 W-133 Technical Meeting. The meeting also benefitted from the expertise and participation of scientists from many state and federal agencies including California Fish and Game, the U.S. Department of Agriculture's Economic Research and Forest Services, the U.S. Department of Interior's Fish and Wildlife Service, and the Bureau of Reclamation. In addition, a number of representatives from the nation's top environmental and resource consulting firms attended, some presenting papers at this year's meeting.

This volume is organized around the goals and objectives of the project, but organizing the papers is difficult because of overlapping themes. The last section includes papers that are very important to the methodological work done by W-133 participants, but do not exactly fit one of the objectives. -- I apologize for the lack of consistent pagination in this volume.

On A Personal Note... Any meeting or conference is successful (and fun!) only because of its participants, so I would first like to thank all the people who came and participated in 1999 - listed below. I also want to thank Jerry Fletcher for all his help at this meeting and prior to it, and John Loomis who passed on his knowledge of how to get a meeting like this to work, and who continues to have the funniest little comments to lighten the meetings up. I especially thank Paul Jakus, who helped me to organize this conference and have a lot of fun during it and afterward. Finally, I want to thank Nicki Wieseke for all her help in preparing this volume, and Billye French for administrative support on conference matters.

W. Douglass Shaw, Dept. of Applied Economics & Statistics, University of Nevada, Reno.
June, 1999

P.S. P.F. and J.C. - As far as I can tell, that darn scorpion is still dead!

CONTROLLING FOR CORRELATION ACROSS CHOICE OCCASIONS AND SITES IN A REPEATED MIXED LOGIT MODEL OF RECREATION DEMAND*

by

JOSEPH A. HERRIGES**

Department of Economics, Iowa State University, 260 Heady Hall, Ames, IA 50011-1070

DANIEL J. PHANEUF

Department of Agricultural and Resource Economics, North Carolina State University, Box 8109, Raleigh, NC 27695-8109

Proposed Running Head: Correlation in Repeated Mixed Logit

June 10, 1999

---

*The authors would like to acknowledge the many helpful comments and suggestions made by Catherine Kling and participants in the 1999 W-133 Regional Project meetings in Tuscon, AZ. We especially want to thank Kenneth Train, both for his helpful insights and for providing the mixed logit estimation software developed by Kenneth Train, David Revelt, and Paul Ruud at the University of California, Berkeley. This research was supported in part by the U.S. Environmental Protection Agency and by the Western Regional Research Project W-133. Although the research described in this article has been funded in part by the United States Environmental Protection Agency through R82-5310-010 to the authors, it has not been subject to the Agency's required peer review policy and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. All remaining errors are our own.
**Corresponding author: 260 Heady Hall, Deparment of Economics, Iowa State University, Ames, IA 50011, *jaherrig@iastate.edu*.

Controlling for Correlation Across Choice Occasions and Sites in a Repeated Mixed Logit Model of Recreation Demand

Abstract

The repeated nested logit model (RNL) introduced by Morey, Rowe and Watson [19] provides a utility consistent approach to controlling for the count nature of recreation demand data. However, it requires strong assumptions on cross-site and cross choice occasion correlation patterns. We examine the use of the mixed logit framework (e.g., [17]) in its place to generalize the available correlation patterns. A Monte Carlo experiment is used to illustrate the ability of the repeated mixed logit model (RXL) to capture quite general correlation patterns, and demonstrate its importance with an application to sport fishing in the Wisconsin Great Lakes.

1. INTRODUCTION

The task of modeling recreation demand is complicated by both the count nature of the data and the prevalence of corner solutions (i.e., households typically choose to visit only a subset of the available sites and set the demand for the remaining sites to zero). Over the past few decades, a variety of frameworks have evolved attempting capture these features of demand. For example, the so-called linked model segments the consumer's decision into two components: (a) the discrete choice of site selection for a given trip and (b) the participation decision regarding the number of trips to be taken.[1] Yet, while the linked model is intuitively appealing and relatively easy to estimate, it cannot be derived from an underlying set of preferences (see, e.g. [13], [27]), making the resulting welfare calculations approximations at best. In contrast, the Kuhn-Tucker (KT) model, initially proposed by Hanemann [10] and Wales and Woodland [30], provides a unified utility theoretic framework within which to model both the site selection and participation decision, while controlling for corner solutions.[2] The KT model does not, however, explicitly control for the count nature of recreation demand data and, as yet, has only been estimated using relatively simple functional forms. Of the competing frameworks currently available in the literature, only the repeated nested logit (RNL) model (introduced by Morey, *et al.* [19]) integrates the site selection and participation decisions in a utility theoretic framework *and* controls for the count nature of recreation demand data. The RNL model is not, however, without its drawbacks. Two key assumptions are employed in its development. First, individuals are assumed to decide whether and where to recreate during a fixed number of choice occasions in a season. For example, it is common practice to assume that households face 52 weekly choice occasions during a year. Second, choice occasion decisions are assumed to be independent not only across individuals, but also across choice occasions for the same individual. This latter assumption precludes habit formation or learning on the part of the recreator and it is the relaxation of this assumption that is the focus of this paper.

Recently, Phaneuf. Kling, and Herriges [24] proposed introducing correlation across choice occasions into the RNL framework by adopting the mixed logit (or random parameters logit) specification developed in McFadden and Train [17].[3] The mixed logit model represents a generalization of multinomial logit, introducing additional error components into the preference structure underlying consumer choices. These error components can modify specific parameters of the individual's preference function (resulting in a "random parameters" specification) or they can be used to capture complex correlation patterns across alternatives and/or choice occasions. The purpose of this paper is to examine the potential for the resulting repeated mixed logit (or RXL) model to address criticisms of the RNL model.

The remainder of the paper is divided into five sections. In Section 2 and 3, we provide an overview of the both the standard RNL specification and the RXL generalization, respectively, illustrating how the latter can be used capture correlation across both choice occasions and alternatives. In the RNL model, correlation across alternatives is captured by identifying *a priori* a nesting structure for the available set of alternatives. While the mixed logit approach can be used to mimic a given nesting structure, we demonstrate how it can be used to allow for more general correlation patterns among alternatives and, indeed, test for specific nesting structures. Furthermore, individual specific error components can be used to introduce correlation across choice occasions. A Monte Carlo experiment is employed in Section 4 to illustrate the properties of the RXL model. Section 5 provides an application of the RNL and RXL models to the demand for recreational angling in the Wisconsin Great Lakes region. Finally, conclusions and recommendations for applying the RXL model are provided in Section 6, along with suggestions for future research.

2. REPEATED NESTED LOGIT

The RNL model begins with the assumption that individuals face a fixed number choice occasions (T) during the course of a season, deciding on each choice occasion whether to stay at home or visit one of the M available sites. The conditional indirect utility that individual i receives from visiting site j during choice occasion t is assumed to take the form

$$\begin{aligned} U_{ijt} &= V_{ijt} + \varepsilon_{ijt} \\ &= V(y_i - c_{ij}, \mathbf{q}_{jt}; \beta) + \varepsilon_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T, \end{aligned} \tag{1}$$

whereas the utility associated with staying at home is given by

$$(2)\qquad \begin{aligned} U_{i0t} &= V_{i0t} + \varepsilon_{i0t} \\ &= V_0(y_i;\beta) + \varepsilon_{i0t}, \qquad i = 1\ldots,N; t = 1,\ldots,T, \end{aligned}$$

where y_i denotes individual i's income per choice occasion, c_{ij} denotes the cost for individual i to visit site j, $\mathbf{q}_j$ is a vector of quality attributes for site j, and β is a vector of unknown parameters. The ε_{ijt}'s are random terms used to capture heterogeneity of preferences in the population, with the ε_{ijt}'s treated as known by the individual but unobserved by the analyst. Finally, the individual is assumed to select on each choice occasion that alternative providing the highest level of utility.

The above framework, in fact, applies to a variety of random utility models. The RNL model emerges by imposing further structure on the way in which preferences vary across individuals (and choice occasions). In particular, the vectors $\varepsilon_{i\cdot t} \equiv (\varepsilon_{i0t},\ldots,\varepsilon_{iJt})'$ are assumed to be independent and identically distributed across individuals *and* choice occasions and drawn from a Generalize Extreme Value (GEV) distribution. Thus, each choice occasion is treated as an independent event characterized by a nested logit model. Nested logit is used at the choice occasion level to allow for correlation (nesting) patterns among the alternatives. Alternatives within the same nest are more similar (i.e., better substitutes) than alternatives in different nests. It is typical, for example, to assume that the recreation sites are grouped into a nest separate from the stay at home alternative (i.e., j=0). Figure 1a illustrates the implied nesting structure. The corresponding choice probabilities are then given by:

$$(3)\qquad P_{ijt} = \begin{cases} 1 - Q_{it} & j = 0 \\ Q_{it} P_{ijt|trip} & j = 1,\ldots,M \end{cases}$$

where

(4) $$Q_{it} = \frac{\left(\sum_{k=1}^{M} \exp\left(V_{ikt}/\theta\right)\right)^{\theta}}{\left(\sum_{k=1}^{M} \exp\left(V_{ikt}/\theta\right)\right)^{\theta} + V_{i0t}}$$

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

(5) $$P_{ijt|trip} = \frac{\exp\left(V_{ijt}/\theta\right)}{\sum_{k=1}^{M} \exp\left(V_{ikt}/\theta\right)} \qquad j = 1,\ldots,M$$

denotes the conditional probability that individual *i* chooses to visit site *j* on choice occasion *t given* that they have decided to take a trip. The parameter θ is known as the dissimilarity coefficient and is required to lie in the unit interval (see, eg., [18]). As θ declines towards zero, greater similarity and correlation exists among choice alternatives within the same nest, whereas as θ approaches one the model reduces to multinomial logit, with independence among the utilities associated with the various choice alternatives ($j = 0,\ldots,M$). This two-level nested logit model imposes the following block diagonal structure on the variance-covariance matrix of the underlying error components:

(6) $$\Sigma \equiv \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0M} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M0} & \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix} = \begin{bmatrix} \sigma_{00} & 0 & \cdots & 0 \\ 0 & \sigma^2_{trip} & \cdots & \rho\sigma^2_{trip} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \rho\sigma^2_{trip} & \cdots & \sigma^2_{trip} \end{bmatrix} = \begin{bmatrix} \sigma_{00} & \mathbf{0}'_M \\ \mathbf{0}_M & \Sigma_{trip} \end{bmatrix},$$

where $\sigma_{ii'} \equiv Cov\left(\varepsilon_{ijt}, \varepsilon_{i'jt}\right) \; \forall j,t$, $\mathbf{0}_n$ is an $n \times 1$ vector of zeros,

(7) $$\Sigma_{trip} = \sigma^2_{trip} \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

and $\rho \in [0,1)$ measures the correlation among the various trip alternatives. To understand the implications of this model, consider the simplest case in which $V_{ijt} = V \;\; \forall i,j,t$, so that only the stochastic terms matter. Intuitively, the structure in (6) argues that if an individual prefers to travel to site 1 rather than staying at home (i.e., because $\varepsilon_{i1t} > \varepsilon_{i0t}$) then they are also more likely to prefer site 2 to staying at home since, given the positive correlation between ε_{i1t} and ε_{i2t}, it is more likely that $\varepsilon_{i2t} > \varepsilon_{i0t}$.

More complex correlation patterns among the choice alternatives can be imposed in the RNL model by further dividing the trip nest (i.e., sites 1 through *M*) into sub-nests. For example, Figure 1b illustrates a three-level nested logit model that groups alternatives 1 through M_A into a sub-nest *A* of similar sites and alternatives $M_A + 1$ through *M* into a sub-nest *B* of similar sites. The resulting choice probabilities take the form:

$$(8) \qquad P_{ijt} = \begin{cases} 1 - Q'_{it} & j = 0 \\ Q'_{it} P_{iAt|trip} P_{ijt|iAt} & j = 1, \ldots, M_A \\ Q'_{it} P_{iBt|trip} P_{ijt|iBt} & j = M_A + 1, \ldots, M \end{cases}$$

where Q'_{it} denotes the probability of taking a trip, $P_{iAt|trip}$ denotes the conditional probability of visiting one of the sites in sub-nest *A given* that a trip is take, and $P_{ijt|iAt}$ denotes the conditional probability of visiting site *j* in sub-nest *A given* that the individual chooses to visit sub-nest *A*.[4] As in the case of the two-level nested logit model, this implies a specific block structure to variance-covariance matrix of the underlying error components:

$$(9) \qquad \Sigma \equiv \begin{bmatrix} \sigma_{00} & \mathbf{0}'_{M_A} & \mathbf{0}'_{M_B} \\ \mathbf{0}_{M_A} & \Sigma_A & \Sigma'_{AB} \\ \mathbf{0}_{M_B} & \Sigma_{AB} & \Sigma_B \end{bmatrix},$$

with $M_B \equiv M - M_A$,

$$(10) \qquad \Sigma_k \equiv \sigma_k^2 \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k \\ \rho_k & 1 & \cdots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \cdots & 1 \end{bmatrix}, k = A, B.$$

$\Sigma_{AB} = \sigma_{AB} J$, where $\sigma_{AB} \equiv \rho_{AB}\sigma_A\sigma_B$, $\rho_{AB} = Corr(\varepsilon_{ijt}, \varepsilon_{ij't})\ \forall j \in A, j' \in B$, and J is an $M_B \times M_A$ matrix of ones). Two alternatives within the same sub-nest are typically assumed to be more correlated than alternatives from different nests (i.e., $\rho_A > \rho_{AB} \geq 0$ and $\rho_B > \rho_{AB} \geq 0$).

The intuition for the three-level nesting structure in (9) is more complicated but similar to that for the two-level nest. Once again, if an individual prefers site 1 to staying at home then they are also more likely to prefer one of the other sites to staying at home, since the "trip" ε_{ijt}'s are positively correlated. Moving down the nesting structure, if the individual prefers the first site in sub-nest A (e.g., j=1) to sites in sub-nest B (i.e. $\varepsilon_{i1t} > \varepsilon_{ij't}, j' \in B$), then it is likely that any site in sub-nest A would be preferred to any site in sub-nest B. This is because the values of the sub-nest A's ε_{ijt}'s are more highly correlated ($\rho_A > \rho_{AB}$), reflecting the belief of the analyst that sites in a sub-nest are close substitutes.

Regardless of the chosen nesting structure, the resulting log-likelihood function is then given by:

$$(11) \qquad LL_{RPL}(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{t=1}^{T} I_{ijt} \ln P_{ijt},$$

where $I_{ijt} = 1$ if individual i chose to visit site j on choice occasion t; = 0 otherwise.

3. REPEATED MIXED LOGIT

Similar to the RNL model, the basic repeated mixed logit (RXL) model begins with the specification of conditional indirect utility functions for the various alternatives. The utility received by individual i during choice occasion t from visiting site j is given by

$$(12) \qquad \begin{aligned} \widetilde{U}_{ijt} &= \widetilde{V}_{ijt}(\widetilde{\beta}_{it}) + \widetilde{\varepsilon}_{ijt} \\ &= V(y_i - c_{ij}, \mathbf{q}_{jt}; \widetilde{\beta}_{it}) + \widetilde{\varepsilon}_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T, \end{aligned}$$

whereas the utility associated with staying at home is given by

$$
(13)\qquad \begin{aligned}\tilde{U}_{i0t} &= \tilde{V}_{i0t}\left(\tilde{\beta}_{it}\right) + \tilde{\varepsilon}_{i0t} \\ &= \tilde{V}_0(y_i;\tilde{\beta}_{it}) + \tilde{\varepsilon}_{i0t}, \qquad i = 1\ldots,N; t = 1,\ldots,T.\end{aligned}
$$

The model deviates from the RNL in two respects. First, the additive error terms (i.e., the $\tilde{\varepsilon}_{ijt}$'s) are assumed to be i.i.d. extreme value variates. Second, the parameter vector $\tilde{\beta}_{it}$ is now assumed to be random, rather than fixed, potentially varying both across individuals and choice occasions. As with the RNL model, all of the random components (i.e., the $\tilde{\varepsilon}_{ijt}$'s and $\tilde{\beta}_{it}$'s) are assumed to be known by the individual, but unobserved by the analyst.

Conditional on the parameter vector $\tilde{\beta}_{it}$, the probability of observing that individual i chooses alternative j on choice occasion t follows the standard logit form:

$$
(14)\qquad Y_{ijt}(\tilde{\beta}_{it}) = \frac{\exp\left[\tilde{V}_{ijt}\left(\tilde{\beta}_{it}\right)\right]}{\sum_{k=0}^{M}\exp\left[\tilde{V}_{ikt}\left(\tilde{\beta}_{it}\right)\right]}.
$$

The corresponding unconditional probability, $\tilde{P}_{ijt}(\varphi)$, is obtained by integrating over an assumed probability density function for the $\tilde{\beta}_{it}$'s. Typically, the $\tilde{\beta}_{it}$'s are assumed to be i.i.d., so that

$$
(15)\qquad \tilde{P}_{ijt}(\varphi) = \int Y_{ijt}(\beta) f(\beta|\varphi) d\beta,
$$

where $f(\beta|\varphi)$ is the pdf for β, parameterized by φ. The log-likelihood is then given by

$$
(16)\qquad LL_{RXL}(\varphi) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{t=1}^{T} I_{ijt} \ln \tilde{P}_{ijt}(\varphi).
$$

While the conditional choice probabilities (i.e., the $Y_{ijt}\left(\tilde{\beta}_{it}\right)$'s) are easy to compute, simulation methods are typically required to compute the unconditional probabilities (the $\tilde{P}_{ijt}(\varphi)$'s) in the process of constructing the maximum likelihood estimates of φ.[5]

An important advantage of the RXL specification over its RNL counterpart is that it allows for greater heterogeneity in individual preferences. The RNL model implicitly allows for shifts in utility in terms of the intercepts (through the ε_{ijt}'s), but restricts marginal effects (such as the marginal utility of income) to be the same across individuals. The mixed logit framework relaxes this latter assumption by treating the $\widetilde{\beta}_{it}$'s as random.

There are two key disadvantages of the basic RXL model outlined above. First, unlike the RNL model, it assumes that, for a given individual and choice occasion, alternative specific utilities are uncorrelated (since the $\widetilde{\varepsilon}_{ijt}$'s are i.i.d.). The nesting structure employed in the RNL framework to group similar alternatives is missing. [6] Second, like the RNL model, there is no correlation in utilities across choice occasions for a given individual.[7] In the following two subsections, the RXL model is generalized so as to relax these two restrictions.

3.1. Cross-site correlation

The RNL model imposes a specific correlation (or substitution) pattern across sites on a given choice occasion by nesting similar alternatives, assuming that the error vector $\varepsilon_{i\cdot t}$ is drawn from the appropriate GEV distribution. As Train [29, p. 127] notes, the analogue to a nest emerges in the mixed logit framework when a random dummy variable is introduced to group certain alternatives. For example, the counterpart to the nested logit model in Figure 1a emerges if equation (12) is replaced by

$$
\begin{aligned}
(17)\qquad \widetilde{U}_{ijt} &= \widetilde{V}_{ijt}\left(\widetilde{\beta}_{it}\right) + \delta_{it} + \widetilde{\varepsilon}_{ijt} \\
&= \widetilde{V}_{ijt}\left(\widetilde{\beta}_{it}\right) + \widetilde{\eta}_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T,
\end{aligned}
$$

where $\widetilde{\eta}_{ijt} \equiv \delta_{it} + \widetilde{\varepsilon}_{ijt}$ and δ_{it} is an i.i.d. random variable that is also independent of $\widetilde{\varepsilon}_{ijt}$. The composite error term $\widetilde{\eta}_{ijt}$ has two components, one that is independent across alternatives ($\widetilde{\varepsilon}_{ijt}$) and one that is shared by all of the trip alternatives (δ_{it}).[8] It is the latter term that captures the "similarity" of the *M* trip alternatives. Individuals with a large positive realization of δ_{it} tend to prefer taking some sort of trip on choice occasion *t*, since the corresponding $\widetilde{U}_{ijt}$'s ($j = 1,\ldots, M$) will, *ceteris paribus*, be larger than $\widetilde{U}_{i0t}$. Similarly, when δ_{it} is negative, all of the trip alternatives will be less attractive. The variance-covariance matrix for $\widetilde{\eta}_{ijt}$ has the same structure as in equation (7).

More complex nesting structures can be created by incorporating additional random dummy variables into the model. For example, equation (17) can be replaced by:

$$
(18)\quad \begin{aligned}\widetilde{U}_{ijt} &= \widetilde{V}_{ijt}(\widetilde{\beta}_{it}) + \delta_{it} + \sum_{k=1}^{M} \tau_{it}^{jk} + \widetilde{\varepsilon}_{ijt} \\ &= \widetilde{V}_{ijt}(\widetilde{\beta}_{it}) + \widetilde{\eta}_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T,\end{aligned}
$$

where now $\widetilde{\eta}_{ijt} \equiv \delta_{it} + \sum_{k=1}^{M} \tau_{it}^{jk} + \widetilde{\varepsilon}_{ijt}$ $(i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T)$ and $\tau_{it}^{kj} = \tau_{it}^{jk}$ $\forall j,k$. The τ_{it}^{jk}'s capture pair-wise similarities of sites.

The structure in (18) can be used to mimic a wide variety of nesting structures. To illustrate this, consider the special case in which $M = 4$. Let $\tau_{it}^{jj} \sim N(\bar{\tau}^{jj}, \sigma^2_{\tau(j,j)})$, $\tau_{it}^{jk} \sim N(0, \sigma^2_{\tau(j,k)})$ $(j \neq k)$, and $\delta_{it} \sim N(0, \sigma^2_{\delta})$. The implied variance-covariance matrix is then given by:

$$
(19)\quad \widetilde{\Sigma} \equiv \begin{bmatrix} \widetilde{\sigma}_{00} & \widetilde{\sigma}_{01} & \cdots & \widetilde{\sigma}_{04} \\ \widetilde{\sigma}_{10} & \widetilde{\sigma}_{11} & \cdots & \widetilde{\sigma}_{14} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\sigma}_{40} & \widetilde{\sigma}_{41} & \cdots & \widetilde{\sigma}_{44} \end{bmatrix} = \begin{bmatrix} \widetilde{\sigma}^2 & 0 & \cdots & 0 \\ 0 & \widetilde{\sigma}_{11} & \cdots & \widetilde{\sigma}_{14} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \widetilde{\sigma}_{41} & \cdots & \widetilde{\sigma}_{41} \end{bmatrix} = \begin{bmatrix} \widetilde{\sigma}^2 & \mathbf{0}_4' \\ \mathbf{0}_4 & \widetilde{\Sigma}_{trip} \end{bmatrix},
$$

where $\widetilde{\sigma}^2 \equiv Var(\widetilde{\varepsilon}_{ijt})$, $\widetilde{\sigma}_{jj} \equiv Var(\widetilde{\eta}_{ijt}) = \widetilde{\sigma}^2 + \sigma^2_{\delta} + \sum_{k=1}^{M} \sigma^2_{\tau(j,k)}$, and $\widetilde{\sigma}_{jk} \equiv Cov(\widetilde{\eta}_{ijt}, \widetilde{\eta}_{ikt}) = \sigma^2_{\delta} + \sigma^2_{\tau(j,k)}$.

The nesting structures in Figure 2 are obtained using the following restrictions:

- Figure 2a: 2-level nest {0, (1,2,3,4)}. This nesting structure, analogous to Figure 1a, is obtained by imposing the restrictions:

(20) **R1**: $\sigma_{\tau(j,k)} = 0$ $\forall j,k$,

so that

(21) $$\tilde{\Sigma}_{trip} = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix},$$

where $a \equiv \tilde{\sigma}^2 + \sigma_\delta^2$ and $b \equiv \sigma_\delta^2$. This is the same pattern as in (7).

- Figure 2b: 3-level nest {0, [(1,2),(3,4)]}. This nesting structure, analogous to Figure 1b, is obtained by imposing the restrictions:

(22) $$\textbf{R2: } \sigma_{\tau(j,k)} = \begin{cases} \sigma_\tau & (j,k) \in \{(1,2),(3,4)\} \\ 0 & \text{otherwise} \end{cases},$$

so that

(23) $$\tilde{\Sigma}_{trip} = \begin{bmatrix} a & c & b & b \\ c & a & b & b \\ b & b & a & c \\ b & b & c & a \end{bmatrix},$$

where $a \equiv \tilde{\sigma}^2 + \sigma_\delta^2 + \sigma_\tau^2$, $b \equiv \sigma_\delta^2$, and $c = \sigma_\delta^2 + \sigma_\tau^2 > b$. This is the same pattern as in (9).

- Figure 2c: 3-level nest {0, [(1,2),3,4]}. This nesting structure is similar to the previous nest, except that alternatives 3 and 4 are not isolated into a separate sub-nest:

(24) $$\textbf{R3: } \sigma_{\tau(j,k)} = 0 \ \forall (j,k) \neq (1,2),$$

so that

(25) $$\tilde{\Sigma}_{trip} = \begin{bmatrix} a & c & b & b \\ c & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}.$$

As the above examples illustrate, a wide variety of "nesting" patterns can be allowed for in the RXL framework. The pairwise nesting suggested above is just one alternative. Other possibilities would include grouping alternatives by type of recreation (e.g., shore versus boat fishing) or by geographical proximity. Furthermore, in contrast to nested logit, it is relatively straightforward to test competing nesting structures when using mixed logit. As Herriges and Kling [12] note, it is common practice in the literature to simply assume a nesting structure when employing nested logit models, without formally testing it against competing assumptions. At best, informal criteria are used, such as consistency with the utility maximization (e.g., [14]) or likelihood dominance (e.g., [12]). Yet, the assumed nesting structure can have a potentially significant impact on the resulting welfare measures ([12]). The mixed logit framework avoids this problem by allowing for *overlapping* nests. For example, after estimating the general specification in (18), the three-level nest {0, [(1,2), (3,4)]} can be explicitly tested for using the restrictions R2 in equation (22). The competing nesting structure {0, [(1,3), (2,4)]} can likewise be tested using the restrictions:

$$\text{(26)} \quad \textbf{R4:}\ \sigma_{\tau\{j,k\}} = \begin{cases} \sigma_\tau & \{j,k\} \in \{\{1,3\},\{2,4\}\} \\ 0 & \text{otherwise.} \end{cases}$$

3.2. Cross Choice Occasion Correlation

Unlike the RNL, the RXL is capable of modeling cross choice occasion correlation by including individual specific error components that are constant over time. This is analogous to random effects models used in continuous panel data models. For example, equation (18) can be generalized as:

$$\text{(27)} \quad \begin{aligned} \widetilde{U}_{ijt} &= \widetilde{V}_{ijt}\left(\widetilde{\beta}_{it}\right) + \gamma_{ij} + \delta_{it} + \sum_{k=1}^{M} \tau_{it}^{jk} + \widetilde{\varepsilon}_{ijt} \\ &= \widetilde{V}_{ijt}\left(\widetilde{\beta}_{it}\right) + \widetilde{\eta}_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T, \end{aligned}$$

where now $\widetilde{\eta}_{ijt} \equiv \gamma_{ij} + \delta_{it} + \sum_{k=1}^{M} \tau_{it}^{jk} + \widetilde{\varepsilon}_{ijt}$. The random component γ_{ij} can be viewed as the unobserved portion of an individual's site utility that does not vary over time.[9] One might, for example, assume that $\gamma_{ij} \sim N\left(0, \sigma_{\gamma(j)}^2\right)$. With this specification, $Cov(\widetilde{U}_{ijt}, \widetilde{U}_{ijs}) = \sigma_{\gamma(j)}^2 \ \forall t \neq s$.

4. MONTE CARLO EXPERIMENT

As the previous section suggests, the RXL model can be used to mimic the familiar nesting structures embodied in the RNL model of recreation demand. Moreover, unlike RNL, the mixed logit framework allows for explicit testing of competing nesting specifications. The purpose of this section is to illustrate these features of the RXL model through a simple Monte Carlo experiment. Data are generated using a RNL model with $M=4$ and the nesting structure in Figure 2b. We then examine how well the RXL model detects the underlying correlation pattern.

Specifically, we assume that 500 individuals (N=500) have four recreation sites to choose from, in addition to the option of not making a trip (M=4), during the course of 5 choice occasions (T=5). Preferences are generated by the simple conditional utility function:

$$U_{ijt} = -\beta c_{ijt} + \varepsilon_{ijt} \qquad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T, \tag{28}$$

where $\beta = 0.003$ denotes the marginal utility of income. The cost of visiting a specific site are assumed to be fixed over time (i.e., $c_{ijt} = c_{ij}\ \forall t$), drawn from uniform distributions for each individual-choice occasion combination (i.e., $c_{ij} \sim i.i.d.U\{50,90\}$). The error vector $\varepsilon_{i\cdot t}$ is assumed to be i.i.d. over time and across individuals, drawn from a GEV value distribution with the three-level nesting structure depicted in Figure 2b. The dissimilarity coefficient for the upper level nest (i.e., trip versus non-trip) is set at $\theta = 0.5$, while the dissimilarity coefficient for the lower level nests (i.e., in choosing between sub-nest (1,2) versus (3,4)) is $\rho = 0.25$. This structure implies that there is greater similarity between alternatives in the same sub-nest (say 1 and 2) than between alternatives in different sub-nests (say 1 and 3). Likewise, there is greater similarity between any two trip options (say 1 and 4) than between a trip option and staying at home (e.g., 0 and 1). The choice probabilities have the form outlined in equation (8), with $M_A = M_B = 2$, $A = \{1,2\}$, and $B = \{3,4\}$.[10] These choice probabilities were then used to simulate choice outcomes for each of the 500 individuals over five choice occasions, yielding a data set with 2500 independent observations.[11] Two hundred such data sets were generated for this Monte Carlo experiment.

In examining the performance of the RXL model, three specifications were estimated for each of the 200 data sets:[12]

- Unconstrained: This model contains dummy variables relating all possible pairs of sites (i.e. $\tau_{it}^{jk} \sim N\left(0, \sigma^2_{\tau(j,k)}\right), \forall j \neq k; j,k = 1,\ldots,4$), in addition to the outer nest dummy variable $\delta_{it} \sim N\left(0, \sigma^2_{\delta}\right)$. That is,

$$(29) \qquad \widetilde{U}_{ijt} = -\beta c_{ij} + \delta_{it} + \sum_{\substack{k=1 \\ k \neq j}}^{4} \tau_{it}^{jk} + \widetilde{\varepsilon}_{ijt} \qquad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T.$$

- "True": This model contains dummy variables relating sites (1,2), (3,4) and (1,2,3,4), imposing the restrictions that $\sigma_{\tau(j,k)} = 0 \ \forall (j,k) \notin \{(1,2),(3,4)\}$.

- "False": This model contains dummy variables relating site (1,3), (2,4) and (1,2,3,4), imposing the restrictions that $\sigma_{\tau(j,k)} = 0 \ \forall (j,k) \notin \{(1,3),(2,4)\}$.

A typical example of the resulting parameters is provided in Table 1. As expected, the unconstrained RXL model detects the correlation pattern implicit in the generated the data. The correlation among all trips (σ^2_{δ}) is statistically significantly at 1% level. Furthermore, the sub-nest correlations between sites 1 and 2 ($\sigma^2_{\tau(1,2)}$) and sites 3 and 4 ($\sigma^2_{\tau(3,4)}$) are also found to be significant using a 1% critical level. In contrast, the remaining pairwise correlations are generally insignificant, with only $\sigma^2_{\tau(2,4)}$ departing significantly from zero and then only at the 5% level. When the restrictions underlying the true nesting structure are imposed (i.e., $\sigma_{\tau(j,k)} = 0 \ \forall (j,k) \notin \{(1,2),(3,4)\}$), as in the third column of Table 1, the restrictions cannot be rejected at any reasonable confidence level.[13] The remaining parameters for the "True" model are all statistically significant. On the other hand, when we attempt to impose the wrong nesting structure, as in the last column of Table 1, the corresponding restrictions are soundly rejected. A likelihood ratio test statistic is χ^2_4=42.92 with a P-value of less than 0.001. Furthermore, the remaining $\sigma^2_{\tau(1,3)}$ and $\sigma^2_{\tau(2,4)}$ are insignificant.

This same pattern of results in Table 1 emerges in general for the 200 replications of this Monte Carlo experiment. Using a 5% critical level, the "False" model restrictions were rejected in 81% of the replications, whereas the true model restrictions were rejected in only 37% of the cases.[14]

5. APPLICATION

The RXL model is illustrated in this section with an application to recreational angling in the Wisconsin Great Lakes region. Results from a comparable RNL model are provided for comparison purposes.

5.1. Data

Data on angling behavior in the Wisconsin Great Lakes region during the 1989 season were gathered via mail surveys by Richard Bishop and Audrey Lyke at the University of Wisconsin – Madison.[15] The surveys provided detailed information about Wisconsin fishing license holders, including the number and destination of fishing trips to the Wisconsin great lakes region, the distances to each region, the type of angling preferred, and socio-demographic characteristics of the respondents. A total of 487 completed survey were available (i.e., $N = 487$), including responses from 240 individuals who visited at least one of the 22 Great Lakes destinations defined in the survey and 247 who fished only inland waters (non-users from the perspective of the Great Lakes region). We have aggregated the destinations of anglers into four sites:

1. South Lake Michigan
2. North Lake Michigan
3. Green Bay
4. Lake Superior

This aggregation divides the Wisconsin portion of the Great Lakes into geographical zones consistent with Wisconsin Department of Natural Resources' classification of the lake region.

The price of a trip to each of the four sites consists of both the direct cost of getting to the site and the opportunity cost of the travel time. Travel costs were computed based on the round trip cost of travel for the vehicle class, while the opportunity cost of time was computed using one-third of the wage rate. The price of a trip is the sum of these two components.

Three site and household characteristic variables are used in our application: fishing catch rates, toxin levels in fish, and an indicator variable for boat ownership. Catch rates were available for the relevant time period from creel surveys by the Wisconsin Department of Natural Resources. A catch rate index was formed as a weighted average of the catch rates for the four aggressively managed salmon species: lake trout, rainbow trout, Coho salmon, and Chinook salmon. In particular, we formed:

$$(30) \qquad R_j \equiv \sum_k w_k R_{kj} ,$$

where w_k denotes the percentage of anglers indicating that they were fishing for the k^{th} fish species (k = lake trout, etc.) and R_{kj} denotes the catch rate for species k at site j.

Toxin levels in fish were obtained from De Vault *et al.* [8]. Toxins provide a good proxy for overall water quality, and are directly responsible for consumption advisories. Results from [8] were matched on the basis of proximity to the four sites defined above. However, toxin levels at the sites are likely to affect visitation rates only if the individual perceives them to create a safety issue. The Wisconsin angling survey provided information regarding this perception. We used this to form an "effective toxins" variable $E_{ij} = T_j D_i$, where T_j denotes the toxin level at site j and $D_i = 1$ if individual i was concerned about fish toxin levels and $D_i = 0$ otherwise.

5.2. Model Specification

Both the RNL and RXL models require specification of conditional indirect utility functions. For the RNL model we assume that conditional indirect utility that individual i receives from visiting site j during choice occasion t takes the form[16]

$$(31) \qquad U_{ijt} = -\beta_1 c_{ij} + \beta_2 R_{ij} + \beta_3 E_{ij} + \varepsilon_{ijt}, \quad i = 1,\ldots,N; j = 1,\ldots, M; t = 1,\ldots,T,$$

whereas the utility associated with staying at home is given by

$$(32) \qquad U_{i0t} = \beta_0 + \beta_4 B_i + \varepsilon_{i0t}, \quad i = 1\ldots,N; t = 1,\ldots,T,$$

where B_i =0 if individual i owns a boat; =0 otherwise. The error terms, ε_{ijt}'s, are assumed to be independent across individuals and choice occasions, with $\varepsilon_{i\cdot t}$ drawn from a GEV distribution with an assumed nesting structure of $\{0,[(2,3),(1,4)]\}$.[17]

A similar specification was used for the RXL model, with

$$(33) \qquad \widetilde{U}_{ijt} = -\beta_1 c_{ij} + \beta_2 R_{ij} + \beta_3 E_{ij} + \widetilde{\eta}_{ijt} \qquad i = 1,\ldots,N; j = 1,\ldots,M; t = 1,\ldots,T,$$

for visits to a recreational site, whereas

$$(34) \qquad \widetilde{U}_{i0t} = \beta_0 + \beta_4 B_i + \widetilde{\eta}_{i0t}, \quad i = 1\ldots,N; t = 1,\ldots,T,$$

for staying at home.[18] The error terms, $\widetilde{\eta}_{ijt}$'s are assumed to take the form:

$$(35) \qquad \widetilde{\eta}_{ijt} = \begin{cases} \gamma_{ij} + \delta_i + \sum_{\substack{k=1 \\ k \neq j}}^{4} \tau_i^{jk} + \widetilde{\varepsilon}_{ijt} & j = 1,\ldots,4; i = 1,\ldots,N; t = 1,\ldots,T \\ \widetilde{\varepsilon}_{ijt} + \gamma_{ij} & j = 0; i = 1,\ldots,N; t = 1,\ldots,T, \end{cases}$$

where $\gamma_{ij} \sim iid\, N(0,\sigma_\gamma^2)$, $\delta_i \sim iid\, N(0,\sigma_\delta^2)$, $\tau_i^{jk} \sim iid\, N(0,\sigma_{\tau(j,k)}^2)$, and $\widetilde{\varepsilon}_{ijt}$ drawn from an extreme value distribution. This specification assumes that the nesting structure, captured by the terms δ_i and τ_i^{jk}, remains constant across choice occasions. Thus, unlike the RNL model, there is correlation across choice occasions (i.e., when $t \neq t'$) for the same individual, since

$$(36) \qquad Cov(\widetilde{\eta}_{ijt}, \widetilde{\eta}_{ij't'}) = \begin{cases} \sigma_\gamma^2 + \sigma_\delta^2 + \sum_{\substack{k=1 \\ k \neq j}}^{4} \sigma_{\tau(j,k)}^2 & j = j' \\ \sigma_\delta^2 + \sigma_{\tau(j,j')}^2 & j \neq j' \end{cases}$$

In the application section below, we estimate both the unconstrained specification and three restricted versions of the model:

- RXL-A: No nesting structure. This model allows cross equation correlations (through γ_{ij}), but allows for no nesting structure (i.e., restricting $\sigma_\delta = 0$ and $\sigma_{\tau(j,k)} = 0 \ \forall j,k$).

- RXL-B: Limited cross-choice occasion correlation: This model allows for a general nesting structure, but requires that there are no cross-choice occasion correlations beyond the nesting structure (i.e., $\sigma_\gamma = 0$).

- RXL-C: Analogue to RNL model: This model imposes an error structure analogous to the RNL's nesting structure of $\{0,[(2,3),(1,4)]\}$. In particular, it imposes the restrictions that there is no cross-choice occasion correlation (i.e., $\sigma_\gamma = 0$) beyond the nesting structure and that

(37) $$\sigma_{\tau\{j,k\}} = \begin{cases} \sigma_\tau & \{j,k\} \in \{\{2,3\},\{1,4\}\} \\ 0 & \text{otherwise.} \end{cases}$$

5.3. Results

The parameter estimates for both the RNL and RXL models are provided in Table 2. Beginning with the RNL results in column 2, we find that the parameter estimates have the expected signs and are uniformly significant at a one-percent critical level. The marginal utility of income (β_1) is positive, with a point estimate of 0.003. As expected, a high catch rate significantly increases the utility of a site, while higher effective toxin levels diminishes utility. Owning a boat reduces the probability of staying at home on a given choice occasion, with $\beta_4 = -1.55$. Finally, both the upper- and lower-level dissimilarity coefficients (θ and ρ, respectively) lie in the unit interval (with $0 < \rho < \theta < 1$) and are significantly different from one, indicating that a distinct correlation patterns exist among the alternative sites. In particular, visits to North Lake Michigan (2) and Green Bay (3) are more similar than visits to North Lake Michigan (2) and Lake Superior (4). Similarly, angling trips are more similar to each other than to the "stay at home" option.

Columns 3 through 6 of Table 2 provide the parameter estimates for the various RXL models. Beginning with the unconstrained specification, several results emerge. First, like in the RNL model, the parameters associated with nonstochastic portion of the RXL utility function (i.e., the β_k's) all have the expected signs and are statistically significant. Second, a complex nesting structure appears to exist among the various alternatives. Like RNL, the four site alternatives (j = 1, 2, 3, 4) are found to be correlated, although σ_δ is not significantly different from zero. Also like the RNL model, sites 2 and 3 (1 and 4) are even more correlated, with $\sigma_{\tau\{2,3\}}$ ($\sigma_{\tau\{1,4\}}$) significantly different from zero at a one percent critical level. This is analogous to the fact that ρ is significantly different from 1 in the RNL model. However, unlike the RNL model, these are not the only cross-site correlations that exist. Indeed, each pairwise correlation term ($\sigma_{\tau\{j,k\}}$) is statistically significant at a one percent critical level.

Third, additional correlation (i.e., beyond the nesting structure) exists across choice occasions, as indicated by the fact that the individual specific error component γ_{ij} is significant (with σ_γ significantly different from zero at a one-percent critical level). Fourth and finally, the unconstrained RXL specification yields a substantial reduction (31%) in the log-likelihood function over its RNL counterpart. Obviously, these two models are not nested, so that a likelihood ratio test does not apply. However, using the likelihood dominance criterion of Pollak and Wales [25], the RXL specification would clearly be preferred.

Columns 4, 5 and 6 of Table 2 represent natural restrictions on the more general RXL specifications. Column 4 considers elimination of the general nesting structure that establishes linkages (i.e., correlations) across alternatives, leaving only the correlation across choice occasions. This restriction is soundly rejected, with the corresponding likelihood ratio statistic of $\chi^2_{df=7} = 66$ and a P-value of less than 0.001. Similarly, when the cross-choice occasion correlation is limited by constraining $\sigma_\gamma = 0$, as in column 5, the restriction is rejected, with $\chi^2_{df=1} = 20$ and a P-value of less than 0.001. Finally, specifying the RXL model so that it mimics the correlation structure implicit in the RNL model (as in the final column of Table 2), the remaining parameters are all statistically significant, yet the restriction is rejected, with likelihood ratio statistic of $\chi^2_{df=5} = 570$ and a P-value of less than 0.001.

5.4. Welfare Analysis

The motivation for estimating models of recreation demand is typically to evaluate the welfare effects of changing site characteristics or availability. In this subsection, we consider two hypothetical changes to conditions in the Wisconsin Great Lakes region:

- A 20% reduction in toxins at each of the four sites,
- Loss of the South Lake Michigan site.

For each scenario, mean compensating variation is computed for the five models presented in Table 2, comparing and contrasting both across the RNL and RXL frameworks and within the RXL approach given different error specifications.

In standard RNL models, the task of computing the compensating variation (*CV*) associated with a change in site characteristics or in the mix of available sites is straightforward, as closed form equations exist. For the RNL model estimated above,

$$(38) \qquad CV = \frac{T}{\beta_1}\left[IV^1 - IV^0\right],$$

where IV^r is the mean inclusive value associated with conditions *r* (*r* = 0 for initial conditions; = 1 for final conditions), with

$$(39) \qquad IV_i^r = \tfrac{1}{N}\sum_{i=1}^{N} \ln\left\{\left[\left(\exp(V_{i2}^r/\rho) + \exp(V_{i3}^r/\rho)\right)^{\rho/\theta} + \left(\exp(V_{i1}^r/\rho) + \exp(V_{i4}^r/\rho)\right)^{\rho/\theta}\right]^{\theta} + \exp(V_{i0}^r)\right\}$$

and

$$(40) \qquad V_{ij}^r = \begin{cases} -\beta_1 c_{ij}^r + \beta_2 R_{ij}^r + \beta_3 E_{ij}^r, & r = 0,1; j = 1,2,3,4 \\ \beta_0 + \beta_4 B_i^r & r = 0,1; j = 0. \end{cases}$$

Computing compensating variation for the RXL model is not as clear-cut and depends upon the interpretation of the error components γ_{ij}, δ_i, and τ_i^{jk}. If, as is typically the case in the literature, these components are treated as representing variation in consumer preferences across individuals in the population, then that variation should be accounted for in calculating compensating variation. A randomly selected individual in the population will have an expected *CV* that depends upon γ_{ij}, δ_i, and τ_i^{jk}, given by:

$$(41) \qquad C\widetilde{V}(\gamma_i,\delta_i,\tau_i) = \frac{T}{\beta_1}\left[I\widetilde{V}^1(\gamma_i,\delta_i,\tau_i) - I\widetilde{V}^0(\gamma_i,\delta_i,\tau_i)\right],$$

where $I\widetilde{V}^r(\gamma,\delta,\tau)$ is the mean inclusive value associated with conditions *r* and error components $\gamma \equiv (\gamma_0,\gamma_1,\gamma_2,\gamma_3,\gamma_4)$, δ_i, and $\tau \equiv (\tau^{12},\tau^{13},\ldots,\tau^{34})$

(42) $$IV_i^r(\gamma,\delta,\tau) = \frac{1}{N}\sum_{i=1}^{N}\ln\left\{\sum_{j=0}^{M}\exp\left[\widetilde{V}_{ij}^r(\gamma,\delta,\tau)\right]\right\}$$

and

(43) $$\widetilde{V}_{ij}^r(\gamma,\delta,\tau) = \begin{cases} -\beta_1 c_{ij}^r + \beta_2 R_{ij}^r + \beta_3 E_{ij}^r + \gamma_j + \delta + \sum_{\substack{k=1\\k\neq j}}^{4} \tau^{jk}, & r = 0,1; j = 1,2,3,4 \\ \beta_0 + \beta_4 B_i^r + \gamma_j & r = 01,; j = 0. \end{cases}$$

The unconditional compensating variation is then constructed using numerical integration, with

(44) $$C\widetilde{V} = \frac{1}{S}\sum_{i=1}^{n} C\widetilde{V}(\gamma^s,\delta^s,\tau^s)$$

where the superscript s is used to denote the s^{th} draw ($s = 1,\ldots, S$) from the estimated distributions for γ, δ, and τ.

An alternative compensating variation results if the error components γ_{ij}, δ_i, and τ_i^{jk} are interpreted as capturing measurement error. In this case, our best estimate of the underlying preference structure for any one individual corresponds to setting the error components to zero. The appropriate welfare measure then becomes

(45) $$CV^* \equiv C\widetilde{V}(0,0,0).$$

In general, it will not be the case that $C\widetilde{V} = CV^*$.[19]

Given the parameter estimates in Table 2, compensating variation estimates are provided in Table 3 for the two scenarios. The columns labeled "Calculation A" use $C\widetilde{V}$ for the RXL models, whereas those labeled "Calculation B" use CV^*. Several results emerge. First, compensating variation estimates vary between the RXL and RNL frameworks and, to a lesser extent, across the error specifications used for the RXL model. For example, using $C\widetilde{V}$ for the RXL models, the compensating variation associated with a twenty- percent reduction in toxin levels differs by almost a factor of two between the RNL ($22) and unconstrained RXL ($41) specifications. The RXL estimates themselves range from $36 under the RXL-B model to $46 under the RXL-C specification.

Similar patterns emerge in the loss of South Lake Michigan scenario.

Second, the interpretation of the error components significantly alters the implied compensating variation. This is particularly true when the error components are used to capture cross-site correlations. For the unconstrained RXL model $C\widetilde{V}$ is over six times CV^*. Unfortunately, there is no observable basis for choosing between these two interpretations.

6. SUMMARY AND CONCLUSIONS

The mixed logit framework has recently garnered considerable attention in the literature, providing a mechanism for generalizing the variety and complexity of the error structures that can be practically built into discrete choice models. Our goal is writing this paper was to explore how this framework can be specifically used to address concerns with the repeated nested logit model of recreation demand. First, can it be used to both test for specific nesting structures and allow for a wider range of nests? Second, can it be used to relax the implicit assumption that individual choices are independent across choice occasions? The answer to both questions appears to be "yes". The Monte Carlo exercise indicates that the RXL model can identify the nesting structure implicit in an RNL model and can in general identify correlation patterns contained in site visitation data. The application indicates that more complex and more general correlation patterns exist in practice than is typically assumed in nested logit models. Furthermore, these correlation patterns matter in terms of the implied welfare effects from changes in site attributes. The correlation across choice occasion also appears to be a significant factor in recreation demand, both in terms of fitted choice probabilities and implied welfare effects.

This research also raises some issues regarding the application of the RXL framework. First, as one might expect, careful specification of the error components is important. Ignoring both correlations across sites and/or across choice occasions can significantly alter the estimated welfare measures. The error components employed here are by no means exhaustive. Additional research is needed into the specification process. Second, and perhaps of greater concern, is the fact that the welfare measures themselves depend upon our interpretation of their source. Traditionally, they have been treated as representing real variation in consumer preferences in the population.[20] However, if they in fact stem from measurement error, then the appropriate welfare calculations can be quite different. Unfortunately, neither the theory nor the data are likely to provide much guidance in choosing between these two interpretations. Instead, analysts should probably compute both measures, with the hope that they do not differ substantially in practice.

7. REFERENCES

1. M. Ben-Akiva, D. Bolduc, and M. Bradley, Estimation of travel choice models with randomly distributed values of time, *Transportation Research Record,* **1413**, 88-97 (1993).
2. C. Bhat, Accommodating variations in responsiveness to level-of-service measures in travel model choice modeling, working paper, Department of Civil Engineering, University of Massachusetts at Amherst (1996).
3. N. E. Bockstael, W. M. Hanemann, and C. L. Kling, Estimating the value of water quality improvements in a recreation demand framework, *Water Resources Research*, **23**, 951-60 (1987).
4. N. E. Bockstael, 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 Protections Agency, Office of Policy Analysis (1986).
5. N. E. Bockstael, and I. E. Strand, The effect of common sources of regression error on benefit estimates, *Land Economics*, **63**, 11-20 (1987).
6. W. Breffle and E. R. Morey, Investigating heterogeneity of preferences in a repeated logit recreation demand model using RP data, working paper, presented at the Western Regional Project (W-133) Annual Meetings, Tucson, AZ., (1999).
7. D. Brownstone and K. E. Train, Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, **89**, 109-129 (1999).
8. D. S. De Vault, D. Dunn, P. Bergqvist, K. Wiberg, and C. Rappe, Polychlorinated dibenzofurans and polychlorinated dibenzo-p-dioxins in Great Lakes fish: A baseline and interlake comparison, *Environmental Toxicology and Chemistry,* **8**, 1013-1022 (1989).
9. P. Feather, D. Hellerstein, and T. Tomasi, A discrete-count model of recreation demand, *J. Environ. Econom. Management*, **29**, 316-22 (1995).

10. W. M. Hanemann, "A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement," Ph.D. dissertation, Department of Economics, Harvard University (1978).

11. J. A. Hausman, G. K. Leonard, and D. McFadden, A utility-consistent, combined discrete choice and count data model: Assessing recreation use losses due to natural resource damage, *J. Public Economics*, **56**, 1-30 (1995).

12. J. A. Herriges and C. L. Kling, The performance of nested logit models when welfare estimation is the goal," *Amer. J. Agr. Econ*, **79**, 792-802 (1997).

13. J. A. Herriges, C. L. Kling, and D. J. Phaneuf, Corner solution models in recreation demand: a comparison of competing frameworks. *in* "Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice" (J. A. Herriges and C. L. Kling, eds.) Edward Elgar Publishing Ltd., Aldershot (1999).

14. C. L. Kling and J. A. Herriges, An empirical investigation of the consistency of nested logit models with utility maximization, *Amer. J. Agr. Econ*, 77(1995): 875-884.

15. A. J. Lyke, *Discrete Choice Models to Value Changes in Environmental Quality: A Great Lakes Case Study*, Ph.D. dissertation, Department of Agricultural Economics, University of Wisconsin-Madison, 1993.

16. D. McFadden, Econometric models of probabilistic choice, *in* "Structural Analysis of Discrete Choice Data with Econometric Applications" (C. F. Manski, and D. McFadden, eds.) The MIT Press, Cambridge (1981).

17. D. McFadden and K. E. Train, Mixed MNL models for discrete response, working paper, Department of Economics, University of California, Berkeley, 1997.

18. E. R. Morey, 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" (J. A. Herriges and C. L. Kling, eds.) Edward Elgar Publishing Ltd., Aldershot (1999).

19. E. R. Morey, R. D. Rowe, and M. Watson, A repeated nested-logit model of Atlantic salmon fishing, *Amer. J. Agr. Econ.* **75**, 578-592 (1993).

20. J. Murdoch, Aggregation bias in conditional and random parameters logit travel-cost models, working paper, presented at the Western Regional Project (W-133) Annual Meetings, Tucson, AZ., (1999).

21. G. R. Parsons and M. J. Kealy, A demand theory for numbers of trips in a random utility model of recreation, *J. Environ. Econom. Management*, **29**, 418-33.

22. D. J. Phaneuf, *Generalized Corner Solution Models in Recreation Demand*, Ph. D. dissertation, Department of Economics, Iowa State University, 1997.

23. D. J. Phaneuf, C. L. Kling, and J. A. Herriges, Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand, *Review of Economics and Statistics*, forthcoming (1999).

24. D. J. Phaneuf, C. L. Kling, and J. A. Herriges, Valuing water quality improvements using revealed preference methods when corner solutions are present, *Amer. J. Agr. Econ.* **80**, 1025-1031 (1998).

25. R. Pollak and T. Wales, The likelihood dominance criterion, *Journal of Econometrics* **47**, 227-242 (1991).

26. D. Revelt and K. E. Train, Mixed logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, **80**, 647-657, (1998).

27. V. K. Smith, Combining discrete choice and count data models: A comment, mimeograph (1997).

28. K. E. Train. Recreation demand models with taste differences over people, *Land Economics* **74**, 230-39 (1998).

29. .K. E. Train, Mixed logit models for recreation demand. *in* "Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice" (J. A. Herriges and C. L. Kling, eds.) Edward Elgar Publishing Ltd., Aldershot (1999).

30. T. J. Wales and A. D. Woodland, Estimation of consumer demand systems with binding nonnegativity constraints, *Journal of Econometrics*, **21**, 263-85 (1983).

Table 1: Example Results for Monte Carlo Experiment[a]

Parameter	Model: Unconstrained	"Truth"	"False"
β	0.011 (0.006)	0.012** (0.001)	0.008** (<0.001)
σ_{δ}	4.06** (0.97)	3.42** (0.74)	1.97** (0.49)
$\sigma_{\tau\{1,2\}}$	2.24** (0.48)	1.87** (0.32)	
$\sigma_{\tau\{1,3\}}$	0.57 (0.41)		0.40 (0.35)
$\sigma_{\tau\{1,4\}}$	0.35 (0.42)		
$\sigma_{\tau\{2,3\}}$	0.35 (0.40)		
$\sigma_{\tau\{2,4\}}$	1.04* (0.41)		0.08 (0.48)
$\sigma_{\tau\{3,4\}}$	2.99** (0.56)	2.52** (0.34)	
Likelihood	-3298.34	-3300.88	-3319.80

[a]Standard Deviations are given in parentheses.

*Significantly different from zero at a 5% level.

**Significantly different from zero at a 1% level.

Table 2: Application

Parameter	Model: RNL	Unconstrained	RXL-A No Nesting	RXL-B Limited Corr.	RXL-C Limited Corr. RNL Nesting
Intercept	2.94**	8.31**	8.66**	8.30**	6.96**
	(0.01)	(0.40)	(0.40)	(0.33)	(0.26)
Income	0.003**	0.008**	0.007**	0.007**	0.012**
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Catch	1.90**	17.68**	19.40**	17.90**	13.38**
	(0.07)	(1.44)	(1.44)	(1.27)	(0.24)
Toxin	-0.04**	-0.12*	-0.13**	-0.105*	-0.19**
	(0.00)	(0.06)	(0.04)	(0.04)	(0.02)
Boat	-1.55**	-2.73**	-3.26**	-2.99**	-2.64**
	(0.01)	(0.32)	(0.27)	(0.30)	(0.40)
θ	0.24**				
	(0.01)				
ρ	0.18**				
	(0.01)				
σ_δ		0.22		0.49**	1.18**
		(0.18)		(0.16)	(0.26)
$\sigma_{\tau\{1,2\}}$		1.22**		1.37**	
		(0.22)		(0.16)	
$\sigma_{\tau\{1,3\}}$		1.73**		1.73**	
		(0.24)		(0.17)	
$\sigma_{\tau\{1,4\}}$		2.24**		2.39**	3.22**
		(0.21)		(0.17)	(0.12)
$\sigma_{\tau\{2,3\}}$		2.64**		2.48**	3.22**
		(0.19)		(0.18)	(0.12)
$\sigma_{\tau\{2,4\}}$		1.10**		1.45**	
		(0.19)		(0.14)	
$\sigma_{\tau\{3,4\}}$		1.34**		0.88**	
		(0.18)		(0.14)	
σ_γ		0.73**	2.44**		
		(0.13)	(0.10)		
Log-likelihood	-8229	-5667	-5700	-5677	-5952

Table 3: Seasonal Welfare Gains

	20% Reduction in Toxins		Loss of South Lake Michigan	
	Calculation A[a]	Calculation B[b]	Calculation A	Calculation B
RNL	\$22.30	\$22.30	-\$322.50	-\$322.50
RXL Unconstrained	\$40.72	\$6.37	-\$637.07	-\$132.70
RXL – A: No Nesting	\$45.67	\$9.26	-\$751.33	-\$200.66
RXL – B: Limited Correlation	\$35.99	\$6.84	-\$710.64	-\$160.06
RXL – C: Limited Correlation & RNL Nesting	\$41.10	\$4.64	-\$398.20	-\$50.69

[a]Mean welfare estimates calculated using repeated draws from estimated parameter distributions.
[b]Mean welfare estimates calculated using means of estimated parameter distributions.

8. FOOTNOTES

[1] The linked model was originally developed by Bockstael, Hanemann, and Strand [4] and Bockstael, Hanemann and Kling [1], with subsequent modifications and applications by Hausman, Leonard, and McFadden [11], Feather, Hellerstein, and Tomasi [9], and Parsons and Kealy [21], among others. See Herriges, Kling and Phaneuf [13] for further discusion.

[2] See Herriges, Kling and Phaneuf [13] and Phaneuf, Kling and Herriges [23] for recent applications to the recreation demand literature.

[3] There have been numerous applications of mixed logit model appearing in the literature as of late, including Ben-Akiv, Bolduc and Bradley [1], Bhat [2], Brownstone and Train [7], Revelt and Train [26], and Train [28,29].

[4] For the sake of brevity, the exact forms of the choice probabilities in equation (7) are not reported here, but are available from the authors upon request.

[5] Descriptions of simulation methods for use with the mixed logit model can be found in [7], [17], and [29], among others. Gauss code incorporating these simulation procedures into a program to estimate mixed logit models (developed by Kenneth Train, David Revelt, and Paul Ruud at the University of California, Berkeley) can be found on Train's home page at http://elsa.berkeley.edu/~train. A modified version of Train's code used in the application section of this paper (for faster estimation in repeated choice situations when characteristics of individuals do not change over time) is available from the authors upon request.

[6] It is important to note, however, that despite the use of i.i.d. extreme value variates for the error term ε_{ijt} ,the basic RXL model does not suffer from the "independence of irrelevant alternatives assumption" that plagues standard logit models. See Train [28,29].

[7] It should be emphasized that throughout this paper when we speak of correlation it is from the perspective of the analyst. As noted above, for the consumer all of the error components are assumed to be known and hence $\widetilde{U}_{ijt}$ is nonstochastic.

[8] See Train [29, pp. 126-128] for additional discussion regarding the interpretation of mixed logit as an error-components model.

[9] Correlation across choice occasions can also be imposed by restricting $\tilde{\beta}_{it} = \tilde{\beta}_i \; \forall t$, as suggested, e.g., in [26], [28], and [29].

[10] The exact equations for the choice probabilities are left to an appendix, available from the authors upon request.

[11] Specifically, for a given individual and choice occasion, the choice probabilities $\{P_{i0t}, P_{i1t}, \ldots, P_{i4t}\}$ can be viewed as dividing the unit interval into 5 segments. A uniform random number generator was then used to select one of these segments and, hence, a specific choice alternative.

[12] The estimation itself was carried out using Gauss and the cross-sectional mixed logit code developed by Train, Revelt, and Ruud (See footnote 5). 100 replications were used in simulating the unconditional choice probabilities. The τ_{it}^{jk}'s and δ_{it} were treated as normally distributed in the mixed logit model, with their means constrained to be zero. The marginal utility of income coefficient β was treated as fixed.

[13] The likelihood ratio test statistic is χ_4^2=5.09 with a P-value of 0.27.

[14] While it is tempting to expect the former percent to be larger and the latter percentage to be smaller, it should be kept in mind that the RXL model provides only an *analogue* to the RNL, mimicking the desired correlation pattern. However, the shape of the underlying distributions are different.

[15] Details of the survey procedures and samples are provided in Lyke [15] and Phaneuf [22].

[16] β_1 denotes the marginal utility of income. Since the conditional utility functions are assumed to be linear in income, the household's base income level becomes irrelevant and is dropped for convenience.

[17] This nesting structure was chosen based upon prior studies using the same data, including [13] and [24].

[18] While the parameters β_k can be specified as random in the mixed logit framework, we have chosen to leave them as nonstochastic in this analysis so as to focus on the cross-site and cross-choice occasion correlations.

[19] This concern about the interpretation of the error term and its effect on welfare calculations is analogous to the concerns raised by Bockstael and Strand [5] in the continuous demand system setting. Previous applications of the mixed logit model in recreation demand (e.g. [28]) have in essence employed the measurement error interpretation in computing welfare effects.

[20] Furthermore, it is what implicitly underlies the CV calculations in equations (39-41) for RNL models.