TWO RUMs unCLOAKED:

Nested-Logit Models of Site Choice

and

Nested-Logit Models of Participation and Site Choice[1]

Edward R. Morey
Department of Economics
Campus Box 256
University of Colorado
Boulder, Colorado 80309-0256
Edward. Morey@Colorado.edu
<http://spot.Colorado.edu/~morey/index.htm>

December 16, 1997

Forthcoming as Chapter 4 in "Valuing the Environment Using Recreation Demand Models," (C.L. Kling and J.A. Herriges, Eds.), Edward Elgar Publishing Ltd.

An earlier version was presented at the W-133 Meetings, Tucson Arizona, February, 1994.

ABSTRACT

Nested logit is increasingly advocated as a tool of recreational demand and benefit estimation. The intent of this short monograph is to lay out, in a simple fashion, the theory behind the nested-logit model of site choice and the nested-logit model of participation and site choice. Rigorous but straightforward derivations of the properties of nested-logit models are provided, including the probability of choosing a particular alternative, likelihood functions, expected maximum utility, and compensating and equivalent variations. Also discussed are the properties of the underlying distribution, estimation, regularity conditions, the interpretation of the scaling parameters, and the relation between those scaling parameters and the Independence of Irrelevant Alternatives (IIA) assumption(s) embedded in both the nested logit model and its special cases. Examples are used to link the theory to recreational demand and benefit estimation.

Those familiar with the 1994 version of this paper will find a few corrections, many more references, and much more elaboration, particularly with respect to the extreme value distribution, regularity conditions, parameter estimation, and consumer's surplus estimation.

---

[1]"Thanks" George, Trudy, Cathy, Sally, Michael, Michael, Joe, Dan, Ray, Bill, Don, Karen, Alpna, Miles, Chris, and Maggie, who snored at my feet.

Policy analysts often require the consumer's surplus associated with a change in the costs or characteristics of a group of consumption activities. The consumer's choice of consumption activity generally involves two simultaneous decisions: whether to participate in a given class of activities and, if so, which alternative to choose from that class. For example, one simultaneously decides both whether to participate in a given class of site-specific recreational activities, and if so, which site to visit. Joint decisions of this type can be modeled in either a multinomial logit (ML) framework or a nested-logit (NL) framework.[2]

Use of the NL model, in contrast to the ML model, is increasingly advocated, particularly when the intent is to model simultaneously both the decision to participate and the choice of site.[3] The argument is that the Independence of Irrelevant Alternatives (IIA) assumption, implicit in the ML model, although often reasonable when all the alternatives are recreational sites of a particular type, can be unreasonable when the sites differ by type or one of the alternatives is nonparticipation. Participation and site choice should therefore be modeled as a two or more stage nested decision that does not impose IIA a priori across all pairs of alternatives. For example, stage one models the participation decision, and stage two models the choice of site given participation. The individual makes the participation and site choice decisions simultaneously. It is common, but unnecessary, to assume that NL models require a sequential decision process.

---

[2]Joint decisions of this type can also be modeled in other frameworks, but these other frameworks are not the topic of this paper.

[3]See Bockstael *et al.* (1986, 1987 and 1991), Carson *et al.* (1987),, Morey *et al.* (1993, 1995 and 1997), Kling and Thomson (1996), and Hoehn *et al.* (1996). Additional examples of discrete-choice models of recreational demand between 1988 and 1997 are Bockstael *et al.* (1989), Creel and Loomis (1992), Jones and Sung (1992), Hausman *et al.* (1995), Milon (1988), Morey *et al.* (1991), Parsons and Kealy (1992 and 1995), and Parsons and Needleman (1992). Earlier examples are Caulkins *et al.* (1986), Feenburg and Mills (1980), Hanemann (1978), and Morey (1981).

The intent of this monograph is to lay out in a simple fashion the NL model and then provide rigorous, but straightforward, derivations of its properties.[4] My motivation for creation of this monograph was to increase my own understanding and experience the pleasures inherent in that process.

Section I derives the probability of choosing an alternative, and then uses it to form some sample-specific likelihood functions. Section II interprets the parameters in the NL model as they relate to unobserved attributes of the alternatives, and the dependence, or independence, of the random components of utility. In this framework, the IIA assumption is discussed. This section also identifies and discusses special cases of NL. Section III advocates Full Information Maximum Likelihood (FIML) estimation. Section IV derives expected maximum utility, Section V discusses budget exhaustion and the other required regularity conditions for conditional indirect utility functions, Section VI derives compensating and equivalent variations, and Section VII expands the nest to three levels. *Example Boxes* are used throughout to link the theory to the application of recreation demand and benefit estimation. *Diversion Boxes* are footnotes that have outgrown the genre.

I. The Two-Level NL Model of Recreational Demand: Its CDF, Probabilities, and Likelihood Function

The intent of this section is to use the basics of probability theory to derive the probability of choosing each alternative from the assumptions that form the basis of the two-level NL model of consumer demand. Once derived, these probability equations can be used to form likelihood functions, the specific form of which depends on the properties of one's sample.

---

[4]Those familiar with the 1994 version of this paper will find a few corrections, many more references, and much more elaboration, particularly with respect to the extreme value distribution, regularity conditions, parameter estimation, and consumer's surplus estimation.

The two-level NL model is designed to explain an individual's choice of alternative when there is a two-dimensional choice set from which the individual **must** choose one of C distinct alternatives, where one of the dimensions of the choice set can be characterized in terms of M distinct types, and the other dimension in terms of J distinct types: $C \leq M \times J$. The individual chooses an alternative, ni, where $n \in M$ and $i \in J$, subject to the restriction that their choice of type in terms of the Jth dimension is consistent with their choice of type in terms of the Mth dimension. Without loss of generality, nest the two dimensions such that if the individual chooses an alternative of type $n \in M$, then the individual's choice of alternative in the J dimension is restricted to a subset of the J types. This subset has J_n elements, where J_m is the number of J types consistent with a choice of type $m \in M$.

Examples: Consider two different two-dimensional models of recreational demand: a model of participation and site choice, and a model of site choice only, but where the sites are of three distinct types.

A model of participation and site choice: Consider a choice set with C=10 alternatives: visiting one of eight fishing sites, staying at home, going bowling, and all other activities. In this case, one might assume M has two elements; 1= fishing and 2=not fishing; J_1=8 (the number of sites) and J_2=2 (staying home, bowling and all other nonfishing activities).

A model of site choice with saltwater sites, lakes, and rivers: Consider a choice set with C=12 alternatives: three saltwater sites, four lakes, and five rivers. In this case, one might assume that M has three elements; 1= saltwater fishing, 2=lake fishing; and 3=river fishing; J_1=3, J_2=4, and J_3=5.

NL models assume the utility the individual receives if he chooses alternative mj is

$$(1) \quad U_{mj} = V_{mj} + \epsilon_{mj} \qquad \forall \ (mj) \in C,$$

where V_{mj} is the systematic component of utility and ϵ_{mj} is a random component.[5] Both terms are known to the individual, but the ϵ_{mj} are unobserved by the researcher, and so they are random variables from the researcher's perspective. Choice is therefore completely deterministic from the individual's perspective;

[5]Two things should be noted about Equation (1). First, it is quite general in that it requires only that V_{mj} exist; that is, the variables that determine V_{mj} are separable from ϵ_{mj}. Second, there are discrete-choice models that fulfill Equation (1) that are not NL models, e.g., multivariate Probit models.

simply put, each time a choice is made, the individual chooses the alternative that provides the most utility.

Consider, in general, the change in utility and choice resulting from a change in $<V_{mj}>$. Given optimal choice in the new state, the choice in the initial state minimizes the utility change resulting from the change in $<V_{mj}>$. If it did not, that alternative would not be maximizing utility in the initial state. But, given optimal choice in the initial state, the choice in the new state maximizes (minimizes) the utility change if the new $<V_{mj}>$ is an improvement (deterioration). One maximizes potential gains and minimizes potential losses. Consider an example: alternative 1, 2, and 3 have utility of 5, 10 and 15 in the initial state and 20, 10 and 15 in the new state. The individual will switch from alternative 3 to 1, and the change in maximum utility is 5. Five is the maximum utility gain given that alternative 3 was initially chosen, and the minimum utility gain given that alternative 1 is chosen in the new state. A change in the vector of V_{mj}, $<V_{mj}>$, might or might not lead to the choice of a different alternative. For example, a decrease in V_{lk} will not result in a different choice if the individual is not currently choosing alternative *lk*, and might not even if he or she is.

Alternatively, the researcher cannot say with certainty which alternative an individual will choose; rather, he or she can merely determine the probability that the individual will choose a particular alternative. Let $<\epsilon_{mj}>$ denote the vector of these C random terms; that is $<\epsilon_{mj}> = \{\epsilon_{11},\epsilon_{12},...,\epsilon_{1J_1},\epsilon_{21},\epsilon_{22},...,\epsilon_{2J_2},...,\epsilon_{M1},\epsilon_{M2},...,\epsilon_{MJ_M}\}$. Let $f(<\epsilon_{mj}>)$ denote their joint probability density function (PDF), and let $F(<\epsilon_{mj}>)$ denote their cumulative density function (CDF).

The probability of choosing a particular alternative is derived by noting that

(2)

$$Prob(ni) = Prob[U_{ni} > U_{mj} \;\forall\; mj \neq ni] = Prob[\epsilon_{mj} < V_{ni} - V_{mj} + \epsilon_{ni} \;\forall\; mj \neq ni].$$

Without loss of generality, order the alternatives so that alternative *ni* is the first alternative, i.e., 11.

Therefore

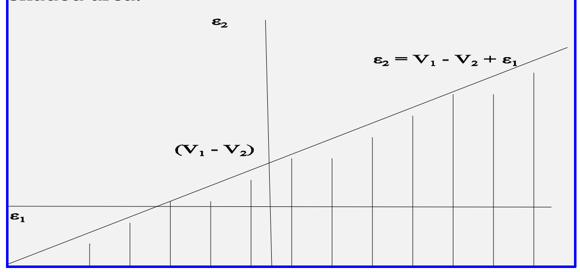
(3) $Prob(11) = Prob[\epsilon_{mj} < V_{11} - V_{mj} + \epsilon_{11} \ \forall \ mj \neq 11]$

$$= \int_{\epsilon_{11}=-\infty}^{+\infty} \int_{\epsilon_{12}=-\infty}^{V_{11}-V_{12}+\epsilon_{11}} \ldots \int_{\epsilon_{mj}=-\infty}^{V_{11}-V_{mj}+\epsilon_{11}} \ldots \int_{\epsilon_{MJ}=-\infty}^{V_{11}-V_{MJ}+\epsilon_{11}} f(\epsilon_{11},\epsilon_{12},\ldots,\epsilon_{mj},\ldots,\epsilon_{MJ}) d\epsilon_{MJ} \ldots d\epsilon_{mj} \ldots d\epsilon_{12} d\epsilon_{11}$$

.

This is the area under the two-dimensional surface, f(<ϵ_{mj}>), where $U_{11} > U_{mj} \ \forall \ mj \neq 11$. Although Equation (3) is a straightforward representation of *Prob(11)*, it can be represented more compactly in terms of the CDF, F(<ϵ_{mj}>). Equation (3) expresses *Prob(11)* as a *C*-level multiple integral; using the CDF, *Prob(11)* can alternatively be expressed as a single integral. The ability to express *Prob(11)*, and more generally *Prob(ni)*, as a single integral makes evaluation of these probability functions much more tractable - who likes to evaluate multiple integrals.

Figure 1: The area under $f(<\epsilon_{mj}>)$ **can be visualized** when *C* has only two elements (e.g., a trip to site 1 or a trip to site 2). $f(<\epsilon_{mj}>) = f(\epsilon_1,\epsilon_2)$ and the probability of visiting site 1, *Prob(1)* is the area under $f(\epsilon_1,\epsilon_2)$ above the shaded area.



The first step in expressing *Prob(11)* in terms of the CDF is to note that, in general,

(4) $Prob[\epsilon_{mj} < \overline{\epsilon_{mj}} \ \forall \ mj \neq 11 : \epsilon_{11} = \overline{\epsilon_{11}}]$

$$= \int_{\epsilon_{12}=-\infty}^{\overline{\epsilon_{12}}} \int_{\epsilon_{mj}=-\infty}^{\overline{\epsilon_{mj}}} \ldots \int_{\epsilon_{MJ}=-\infty}^{\overline{\epsilon_{MJ}}} f(\overline{\epsilon_{11}},\epsilon_{12},\ldots,\epsilon_{mj},\ldots,\epsilon_{MJ}) d\epsilon_{MJ} \ldots d\epsilon_{mj} \ldots d\epsilon_{12}$$

$$= F_{11}(\overline{\epsilon_{11}},\overline{\epsilon_{12}},...,\overline{\epsilon_{mj}},...,\overline{\epsilon_{MJ}}),$$

where $F_{ni}(.)$ denotes the derivative of F with respect to its $(ni)^{th}$ argument, and the *bar over a variable*, $\bar{}$, denotes a specific value of that variable. Equation (4) tells us that the area under the density function defined in the middle term of Equation (4), which is a probability, can be expressed as a derivative of the CDF.[6] The probability that $[\epsilon_{mj} < \overline{\epsilon_{mj}} \ \forall \ mj \neq 11]$ is then obtained by integrating Equation (4) with respect to ϵ_{11} from minus to plus infinity. That is,

(5) $$Prob[\epsilon_{mj} < \overline{\epsilon_{mj}} \ \forall \ mj \neq 11] = \int_{\epsilon_{11}=-\infty}^{+\infty} F_{11}(\epsilon_{11},\overline{\epsilon_{12}},...,\overline{\epsilon_{mj}},...,\overline{\epsilon_{MJ}})d\epsilon_{11}.$$

Utilizing Equations (3) and (5), the probability of choosing alternative *11* is, in terms of the CDF,

(6) $$Prob(11) = Prob[\epsilon_{mj} < V_{11} - V_{mj} + \epsilon_{11} \ \forall \ mj \neq 11]$$

$$= \int_{\epsilon_{11}=-\infty}^{+\infty} F_{11}(<V_{11}-V_{mj}+\epsilon_{11}>)d\epsilon_{11},$$

where $<V_{11}-V_{mj}+\epsilon_{11}> = \{\epsilon_{11}, V_{11}-V_{12}+\epsilon_{11},...,V_{11}-V_{mj}+\epsilon_{11},...,V_{11}-V_{MJ}+\epsilon_{11}\}$. However, since there is nothing unique about alternative 11,

(7) $$Prob(ni) = \int_{\epsilon_{ni}=-\infty}^{+\infty} F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>)d\epsilon_{ni},$$

where $<V_{ni}-V_{mj}+\epsilon_{ni}> = \{V_{ni}-V_{11}+\epsilon_{ni}, V_{ni}-V_{12}+\epsilon_{ni},...,V_{ni}-V_{ni}+\epsilon_{ni},...,V_{ni}-V_{MJ}+\epsilon_{ni}\}$.

As noted above, Equation (7) is preferred over Equation (3) because it is a single integral, whereas Equation (3) is a *C*-level multiple integral. Equation (7) is the probability of choosing alternative *ni* for

---

[6] The second step in Equation (4) follows from the fact that $F(\overline{\epsilon_{11}},\overline{\epsilon_{12}},...,\overline{\epsilon_{mj}},...,\overline{\epsilon_{MJ}}) = \int_{-\infty}^{\overline{\epsilon_{11}}}...\int_{-\infty}^{\overline{\epsilon_{mj}}}...\int_{-\infty}^{\overline{\epsilon_{MJ}}} f(\epsilon_{11},...\epsilon_{mj},...\epsilon_{MJ})d\epsilon_{11},...d\epsilon_{mj},...d\epsilon_{MJ}$. Take the derivative of both sides with respect to ϵ_{11}.

any model that assumes Equation (1). Up to this point the model is very general; it is consistent with any $F(<\epsilon_{mj}>)$.

To generate a two-level NL model, specifically assume that the CDF is[7]

$$(8)\quad F(<\epsilon_{mj}>) = \exp\{-\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{-s_m\epsilon_{mj}}]^{(1/s_m)}\} = \exp\{-\sum_{m=1}^{M} [\sum_{j=1}^{J_m} e^{-s_m(\epsilon_{mj}-\alpha_m)}]^{(1/s_m)}\},$$

where $a_m = e^{\alpha_m}$, $a_m > 0$ *and* $s_m \geq 1\ \forall m$.[8] The condition $(a_m > 0$ *and* $s_m \geq 1\ \forall m)$ is sufficient to guarantee that Equation (8) is a globally well-behaved CDF (see *Diversion 1* for more details). The $<a_m>$ position the distribution and the $<s_m>$ determine its variances and covariances. This CDF is a special case of a multivariate generalized extreme value distribution. The generalized extreme value distribution was first proposed by McFadden (1978). The task at hand is to show that the derivative of the this CDF, when

---

[7]Alternatively, if one assumed a multivariate normal CDF, the model would be multivariate Probit.

[8] Those familiar with McFadden (1978) will note that I have not used his notation. My s_m is his *1/(1 -* σ_m*)*. I find my notation simpler both in terms of wordprocessing and comprehension. Note that $s_m \geq 1 \Leftrightarrow 0 \leq \sigma_m < 1$. Alternatively Börsh-Supan (1990) and Kling and Herriges (1995 and 1996) use the notation θ_m, *where* $s_m = 1/\theta_m$. They refer to θ_m as a *dissimilarity coefficient.*

plugged into Equation (7), generates the *Prob(ni)* equation for the two-level NL model.

A major reason for choosing this particular CDF, Equation (8), is that it generates a closed form solution for Equation (7). This greatly simplifies estimation of the model, eliminating the need for numerical integration. Most CDFs do not generate closed forms for the *Prob(ni)*. For example, if one assumes the CDF is multivariate normal (the multivariate Probit model), the Equation (7) integral will not have a closed- form solution, so estimation of the likelihood function requires complex, numerical, multiple integration. This is why estimated multivariate probit models limit the choice set to a small number of alternatives (e.g.,two, three, four), but NL models can be estimated with large numbers of alternatives.[9]

Diversion 1: **Equation (8) is not a globally well-defined CDF** $\forall$ **$< a_m >$ and $< s_m >$**, and this often causes problems when the $< s_m >$ are estimated.

Simply put, for a function, $F(<\epsilon_{mj}>)$, to be a well-defined CDF its range must be the unit interval and it must be nondecreasing in $<\epsilon_{mj}>$. This second condition can be alternatively stated as $F(<\epsilon_{mj}>)$ must be the integral of a nonnegative density function, $f(<\epsilon_{mj}>)$. Since $f(<\epsilon_{mj}>) = \frac{\partial}{\partial\epsilon_{11}}\frac{\partial}{\partial\epsilon_{12}}...\frac{\partial}{\partial\epsilon_{MJ}}F(<\epsilon_{mj}>)$, the nonnegativity of $f(<\epsilon_{mj}>)$ requires that the mixed partial, $\frac{\partial}{\partial\epsilon_{11}}\frac{\partial}{\partial\epsilon_{12}}...\frac{\partial}{\partial\epsilon_{MJ}}F(<\epsilon_{mj}>)$, is always nonnegative.

Assume $a_m > 0 \;\forall\; m$. Given this, $s_m \geq 1 \;\forall_m$ is necessary and sufficient for Equation (8) to be a well-defined CDF. The condition $(s_m \geq 1 \;\forall_m)$ is typically referred to as the Daly-Zachary-McFadden condition, except that it is often stated $(0 < \theta_m \leq 1 \;\forall m)$. Looking ahead, when $(s_m \geq 1 \;\forall\; m)$ the model is said to be globally consistent with stochastic utility maximization.

Sometimes estimated s_m are less than one. For such cases, Börsh-Supan (1990) and Herriges and Kling (1996) consider whether the estimated model is locally consistent with stochastic utility maximization.

[9] Advances are being made in this area. See, for example, McFadden(1989), Pakes and Pollard (1989), Layton (1996), Chen (1996) and Chen *et al.* (1997).

Examples: Consider the F(<ϵ_{mj}>) for the two models of recreational demand introduced in the first example box.

The model of participation and site choice where *M* has two elements: 1= fishing and 2=not fishing, J_1=8 (the number of sites) and J_2=2 (staying home and bowling). For this example,

$F(<\epsilon_{mj}>)$
$= \exp\{-(a_1[\exp(-s_1\epsilon_{11})+...+\exp(-s_1\epsilon_{18})]^{(1/s1)}$
$+a_2[\exp(-s_2\epsilon_{21})+\exp(-s_2\epsilon_{22})]^{(1/s2)})\}$.

The model of site choice with saltwater sites, lakes and rivers, where *M* has three elements: 1= saltwater fishing, 2=lake fishing, and 3=river fishing; J_1=3, J_2=4 and J_3=5. For this example,

$F(<\epsilon_{mj}>)$
$= \exp\{-(a_1[\exp(-s_1\epsilon_{11})+...+\exp(-s_1\epsilon_{13})]^{(1/s1)}$
$+a_2[\exp(-s_2\epsilon_{21})+...+\exp(-s_2\epsilon_{24})]^{(1/s2)}$
$+a_3[\exp(-s_3\epsilon_{31})+...+\exp(-s_3\epsilon_{35})]^{(1/s3)})\}$.

To obtain the closed form of the *Prob(ni)* equation, first take the derivative of the multivariate extreme value CDF with respect to its $(ni)^{th}$ element. One obtains

$$(9)\quad F_{ni}(<\epsilon_{mj}>) \;=\; \exp\{-\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{-s_m(\epsilon_{mi})}]^{(1/s_m)}\}\; a_n[\sum_{j=1}^{J_n} e^{-s_n\epsilon_{nj}}]^{(1/s_n)-1}\; e^{-s_n\epsilon_{ni}}.$$

Substituting $<V_{ni}+\epsilon_{ni}-V_{mj}>$ for $<\epsilon_{mj}>$ in Equation (9), one obtains

(10) $F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>) = \exp\{-\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{-s_m(V_{ni}+\epsilon_{ni}-V_{mj})}]^{(1/s_m)}\}$

$$\{a_n[\sum_{j=1}^{J_n} e^{-s_n(V_{ni}-V_{nj}+\epsilon_{ni})}]^{(1/s_n)-1} e^{-s_n\epsilon_{ni}}\}.$$

Before substituting the RHS of Equation (10) into Equation (7) to obtain *Prob(ni)*, simplify Equation (10) into terms that do, and do not, involve ϵ_{ni} so that $F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>)$ in Equation (7) will be easy to integrate with respect to ϵ_{ni}. Factoring Equation (10) one obtains

(11) $F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>) = e^{-\epsilon_{ni}} \exp\{-e^{-\epsilon_{ni}}e^{-V_{ni}}B\}\ A,$

where

(12) $A \equiv a_n[\sum_{j=1}^{J_n} e^{s_nV_{nj}}]^{(1/s_n)-1}\ e^{-V_{ni}}\ e^{s_nV_{ni}},$ and

(13) $B \equiv \sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{s_mV_{mj}}]^{1/s_m}.$

Note that A and B do not depend on ϵ_{ni}. Plugging Equation (11) into Equation (7), one obtains

(14) $Prob(ni) = A \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} \exp\{-e^{-\epsilon}\ Z\ B\}\ d\epsilon\ ,$

where $Z \equiv e^{-V_{ni}}$. Rather than trying to integrate this with respect to ϵ, simplify it further by making the change of variables w = $e^{-\epsilon}$ *($\Rightarrow d\epsilon$ = -(1/w)dw)* to obtain

(15)

$Prob(ni) = -A \int_{w=0}^{\infty} \exp\{-wZB\}\ dw = A/(ZB)$

.

Substituting back in for *A*, *B*, and *Z*, one obtains

Example: For the **model of site choice with saltwater sites, lakes and rivers**, the probability that the individual will choose the third lake site (ni = 23) is
Prob(23) =
$\exp(s_2V_{23})a_2[\exp(s_2V_{21})+...+\exp(s_2V_{24})]^{((1/s2)-1)}$
divided by
$\{a_1[\exp(s_1V_{11})+...+\exp(s_1V_{13})]^{1/s1}$

$$(16) \quad Prob(ni) = \frac{e^{s_n V_{ni}} \; a_n [\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1}}{\sum_{m=1}^{M} a_m [\sum_{j=1}^{J_m} e^{s_m V_{mj}}]^{1/s_m}} ,$$

which is the probability of choosing alternative *ni* in a two-level NL model. Assuming that the elements of <a_{mj}> are all positive (see *Diversion 1*), inclusion of the <a_{mj}> parameters is equivalent to adding a group-specific constant term, α_m, to each of the V_{mj}, where $\alpha_m = \ln(a_m)$. To see this, replace a_m with e^{α_m} ($\Rightarrow \alpha_m = \ln(a_m)$); in which case *Prob(ni)* can be rewritten as

(16a) $$Prob(ni) = \frac{e^{s_n(\alpha_n + V_{ni})} [\sum_{j=1}^{J_n} e^{s_n(\alpha_n + V_{nj})}]^{(1/s_n) - 1}}{\sum_{m=1}^{M} [\sum_{j=1}^{J_m} e^{s_m(\alpha_m + V_{mj})}]^{1/s_m}}$$

Diversion 2: Conditions on the *Prob(ni)*

The *Prob(ni)* have the following properties for all $<v_{mj}> \in R^C$: (i) each *Prob(ni)* is greater than zero and less than one; (ii) the *Prob(ni)* sum to one; (iii) $\frac{\partial Prob(lk)}{\partial mj} = \frac{\partial Prob(mj)}{\partial lk}$ and (iv) the *Prob(ni)* are invariant to replacing each α_m *with* $(\alpha + \alpha_m) \; \forall \; \alpha \in R$ (⇔adding a positive constant to each of the a_m). Note that the *Prob(ni)* possess properties (i)-(iv) even if the s_m are not all ≥ 1, so these properties hold even if Equation (8) is not a well-behaved CDF. Given $(a_m > 0 \; \forall \; m)$, $(s_m \geq 1 \; \forall \; m)$ is necessary and sufficient to ensure that these probabilities are consistent with an underlying density function $\forall \; <V_{mj}> \in R^C$. More generally, given probabilities that possess properties (i)-(iv), such probabilities will always be consistent with a well-behaved underlying density function if $\forall$ *mj*, the mixed partial derivative of *Prob(mj)* with respect to the V_{lk}, $lk \neq mj$, is always nonpositive when the mixed partial is of an odd-order and nonnegative when the mixed partial is of an even-order (Herriges and Kling, 1996). For Equation (16a), $(s_m \geq 1 \; \forall \; m)$, and these sign restrictions on the mixed partials are equivalent. As noted in *Diversion 1*, when these conditions are fulfilled, the NL model is said to be globally consistent with stochastic utility maximization.

Alternatively, if some of the s_m are less than 1, it is still possible that, for the $<V_{mj}>$ of interest, the aforementioned mixed partials all have the correct signs; that is, $\forall$ current and proposed levels of the $<V_{mj}>$, the mixed partials have the correct signs. In such cases, the *Prob(ni)* are said to be locally consistent with stochastic utility maximization.

If desired, the probability, Equation (16), can be decomposed into the probability of choosing an alternative of type *n* multiplied by the probability of choosing alternative i from the group of alternatives that are of type *n*; i.e.,

(17) $$Prob(ni) = Prob(i|n)Prob(n), \quad \text{where}$$

(18) $$Prob(n) = \frac{a_n[\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{1/s_n}}{\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{s_m V_{mj}}]^{1/s_m}} \quad \text{and}$$

(19) $$Prob(i|n) = \frac{e^{sV_{ni}}}{[\sum_{j=1}^{J_n} e^{sV_{nj}}]}.$$

Equation (16) is made explicit by specifying functional forms for the $V_{mj,}$ where V_{mj} is the conditional indirect utility function for alternative *mj*. V_{mj} is assumed to be a function, often linear, of the cost of alternative *mj*, the budget, the characteristics of alternative *mj*, and characteristics of the individual. For example, if *mj* is a fishing site, the variables might be the cost of a trip to site *mj*, the expected catch rate at site *mj*, and other characteristics of the site. The regularity conditions on the V_{mj} are considered in Section V.

> **Examples:** Consider what constitutes a **choice occasion** for the two models of recreation demand introduced in the first example box.
>
> In the model of **participation and site choic**e, the fishing season can be divided into a finite number of periods and each period is a choice occasion. For example, each week in the season might be defined as a choice occasion, or, more generally, the season might be divided into a fixed number of periods, but with no restriction that each period is of a specified length, just that the sum of the choice occasions equal the season. The critical issue for this simple model of participation and site choice is that no more than one site is chosen on each choice occasion.
>
> In the model of **site choice with saltwater sites, lakes and rivers**, a choice occasion is each time a fishing trip is taken.

Consider now the problem of estimating the parameters in the V_{mj} functions using a random sample of individuals that reports the alternative, or alternatives, chosen by each individual in the sample. Note that the choices can be actual choices, hypothetical choices, or a combination of both.[10]

At this point it is important to make a distinction. Denote each time an individual must choose between the C alternatives in the choice set as a *choice occasion.* An important distinction is whether the sample contains information on the alternative chosen for just one choice occasion for each individual, or whether the data set reports, by individual, the alternative chosen on each of a number of choice occasions. The number of observed choice occasions could vary across individuals. Start with the simpler case where the sample contains the choice on only one choice occasion for each individual. The probability of observing individual *h* choosing alternative *ni* on the one choice occasion is *Prob(hni)*. If one further assumes that the choices of the *H* individuals in the sample are statistically independent, $\text{cov}(\epsilon_{hni}, \epsilon_{kmj}) = 0$, $\forall\ n,i,m,j,\ h \neq k$, the probability of observing the set of observed choices is

$$(20)\quad f(\langle y_{1mj}\rangle,\langle y_{2mj}\rangle,\ldots,\langle y_{Hmj}\rangle) = \prod_{h=1}^{H}\prod_{n=1}^{M}\prod_{i=1}^{J_n} Prob(hni)^{y_{hni}},$$

where $y_{hni} = 1$ if individual *h* choose alternative *ni*, and zero otherwise, *C* is the number of alternatives in the choice set, *Prob(hni)* is defined by Equation (16), and

$$\langle y_{hmj}\rangle = \{y_{h11},y_{h12},\ldots,y_{h1J_1},y_{h21},y_{h22},\ldots,y_{h2J_2},\ldots,y_{hM1},y_{hM2},\ldots,y_{hMJ_M}\}.$$

Equation (20) is the likelihood function for this sample; that is, it is the probability of observing the choices in the sample as a function of the *Prob(hmj)*. The task is to find those values of the parameters in the *Prob(hmj)* that maximize the likelihood function. Since the parameters that maximize the log of the

---

[10]Examples of ML and NL models of site-specific recreational activities choice estimated using hypothetical choices or hypothetical and actual choices include Adamowicz *et al.* (1994, 1996), Swait and Adamowicz (1996), and Buchanan *et al.* (1997). Morey et al. (1997b) uses choices among pairs of hypothetical alternatives to estimate a ML model for monument preservation programs.

likelihood function also maximize the likelihood function, estimation is simplified by finding those values of the parameters that maximize the log of the likelihood function.

$$(21)\quad L = \sum_{h=1}^{H} \sum_{n=1}^{M} \sum_{i=1}^{J_n} y_{hni} \ln[Prob(hni)].$$

Estimation of log likelihood functions are discussed in section III.

Consider now the case where the data set reports, by individual, the alternative chosen on each of a number of choice occasions, where the number of observed choice occasions may vary across individuals. Such samples are generated by *repeated-choice* problems; that is, the discrete-choice problem faced by each individual repeats, so there are multiple choice occasions. Some discrete choice problems such as what furnace to purchase or what individual to marry do not repeat, or, hopefully, do not repeat often. In contrast, discrete choice problems in recreational demand are characterized by repetition. The problem of where to go on a fishing trip repeats every time one takes a trip. The problem of whether to take a fishing trip also repeats every choice occasion (for example, every day or every week).

> **Examples: The log likelihood functions** for the two models of recreational demand introduced in the first example box when the sample only reports the alternative chosen on one choice occasion for each individual in a sample of H=100 individuals is Equation (21), where for
>
> **The model of participation and site choice** M has two elements: 1= fishing and 2=not fishing; J_1=8 (the number of sites) and J_2=2 (staying home and bowling).
>
> **The model of site choice with saltwater sites, lakes and rivers**, M has three elements: 1= saltwater fishing, 2=lake fishing, and 3=river fishing; J_1=3, J_2=4, and J_3=5.

Let T_h denote the number of choice occasions observed for individual h and ϵ_{hmjt} the random component in the utility individual h receives during choice occasion t if alternative mj is chosen. Assume, in addition to the previous assumption, that choices are statistically independent across individuals, and that choices for a given individual are statistically independent across choice occasions. That is, assume $\mathrm{cov}(\epsilon_{knis}, \epsilon_{kmjt}) = 0,\ \forall\ k,n,i,m,j,\ s \neq t.$[11]

For this case, let $Y_{hni} \equiv$ the number of times individual *h* chooses alternative *ni*, where $\sum_{n=1}^{M} \sum_{i=1}^{J_n} Y_{hni} = T_h$. The probability of observing the vector of alternatives $<Y_{hni}>$ for individual *h* is determined by the multinomial density function.

---

[11]One way to generate correlation across choice occasions is to assume individual-specific, alternative-specific random effects that cannot be attributed to observable variations in alternative and individual characteristics. We have assumed $U_{hmjt} = V_{hmjt} + \epsilon_{hmjt}$, where each $<\epsilon_{hmjt}>$ is a random draw from Equation (8). Now consider replacing ϵ_{hmjt} with the sum of two random components, $\nu_{hmj} + \epsilon_{hmjt}$, where, as above, the $<\epsilon_{hmjt}>$ are independent across choice occasions, but the $<\nu_{hmj}>$ remain the same across choice occasions. One might view $<\nu_{hmj}>$ as randomly drawn once a year from some PDF, $f(<\nu_{hmj}>)$, so that $<\nu_{hmj}>$ varies from year to year but not across choice occasions within a year, or one might assume the individual's $<\nu_{hmj}>$ is drawn only once. The $<\nu_{hmj}>$ add individual-specific, alternative-specific constants to the conditional indirect utility functions, and these terms remain constant across choice occasions, so cause the random term to be correlated across choice occasions. $Prob(hnit \mid <\nu_{hni}>)$ is the standard nested logit probability after these alternative specific constants have been added to the deterministic component of utility for each alternative. Assuming the $<\nu_{hmj}>$ do not depend on the explanatory variables in the model, $Prob(hnit, <\nu_{hmj}>) = Prob(hnit \mid <\nu_{hni}>)\ f(<\nu_{hmj}>)$. Note that this joint probability depends on the unobservable $<\nu_{hmj}>$. Proceed by specifying a functional form for $f(<\nu_{hmj}>)$ such that when one sequentially integrates *Prob(hni)* from $-\infty$ *to* $+\infty$ with respect to each of the elements of $<\nu_{hmj}>$, one obtains a computable form for *Prob(hni)* as a function of the parameters and explanatory variables in the NL model, and the parameters in $f(<\nu_{hmj}>)$. This is not a trivial exercise. Simplifying assumptions include $f(<\nu_{hmj}>) = \prod_{m=1}^{M} \prod_{j=1}^{J_m} f(\nu_{hmj})$ and some or most of the elements of $f(<\nu_{hmj}>)$ are zero (only some alternatives have individual-specific, alternative-specific random effects).

Alternatively, *Diversion 3* outlines a different method to allow correlations across choice occasions, a method that allows the variance of the random components to systematically differ across individuals. However, the method does not allow an individual to have persistent site preferences that are independent of site and individual characteristics included in the model.

$$(22)\quad f(<Y_{hmj}>) \;=\; \frac{T_h!}{\prod_{n=1}^{M}\prod_{i=1}^{J_n} Y_{hni}!}\;\prod_{n=1}^{M}\prod_{i=1}^{J_n}\; Prob(hni)^{Y_{hni}}.$$

The log of the likelihood function for this sample is[12]

$$(23)\quad L \;=\; \sum_{h=1}^{H}\sum_{n=1}^{N}\sum_{i=1}^{J_n}\; Y_{hni}\;\ln[Prob(hni)].$$

Note that it is possible to estimate the number of choice occasions by maximizing Equation (23) for different values of *T*. For an example, see Morey *et al.* (1991).

For example, in our **model of participation and site choice** one might know for some choice occasions both whether and where an individual fished, but for other choice occasions know that a trip was taken but not have information about the destination.

In the **model of site choice for saltwater sites, lakes and rivers**, one might know for some choice occasions the exact site chosen, but for other choice occasions only that the trip was to a river.

Another common but more complicated type of sample is a sample that contains, for each individual, information on the specific alternative chosen for some choice occasions, but only partial information on the alternative chosen for other choice occasions. For example, one might know the specific alternative chosen for some choice occasions but for others only know which of the *M* groups contains the alternative. The log of the likelihood function for such an "incomplete" sample is

$$(24)\quad L \;=\; \sum_{h=1}^{H}\sum_{n=1}^{M}\{Y_{hn}\ln[Prob(hn)] \;+\; \sum_{i=1}^{J_n}\; Y_{hni}\;\ln[Prob(hni)]\}\;,$$

where *Prob(hn)* is defined in Equation (18), Y_{hn} is the number of times we know individual *i* chose an alternative of type *n* but not which one. Note that in this case $T_h \;=\; \sum_{n=1}^{N}\sum_{i=1}^{J_n} Y_{hni} \;+\; \sum_{n=1}^{N} Y_{hn}$.

[12]The additive term, *ln*(multinomial coefficient), is omitted because it does not depend on the values of the parameters in the *Prob(hmj)*.

II. Special Cases of the Two-Level NL Model

The significance of the a_m and s_m parameters in the CDF, Equation (8), is deciphered by remembering that the utility an individual receives if he chooses alternative *ni* is, from the analyst's perspective, a random variable; i.e., $U_{ni} = V_{ni} + \alpha_m + \epsilon_{ni}$, where $\alpha_m = \ln(a_m)$, $(V_{ni} + \alpha_m)$ is deterministic, and ϵ_{ni} is the random variable.[13] The V_{ni} are a function of the *attributes* of the alternatives that are observed by the analyst. The $(\alpha_n + \epsilon_{ni})$ are the impacts of the unobserved attributes. As noted earlier, inclusion of the $<a_m>$ parameters is equivalent to adding a group-specific constant term, α_m, to each of the V_{mj}; this will be elaborated on below.

A critical issue in all two-level discrete choice models is whether each element of the vector $<\epsilon_{mj}>$ is **independently** drawn from the same univariate distribution, or whether elements of $<\epsilon_{mj}>$ are drawn from a multivariate distribution and are therefore correlated. The NL CDF, Equation (8), allows the ϵ_{mj} to be correlated by type.

What would cause the random terms, $<\epsilon_{mj}>$, to be correlated by type? If an attribute that is an important determinant of choice is not observed, it influences the magnitude of the $<\epsilon_{mj}>$, the $<\alpha_m>$, or both. If this attribute **varies** across alternatives within a group less, or more, than across alternatives in different groups, the random elements in group *n* will be more correlated with each other than they are with the

[13]Note that the α_m can be viewed as either a parameter in the CDF, Equation (8), as it is here, or as a possible component of V_{mj}.

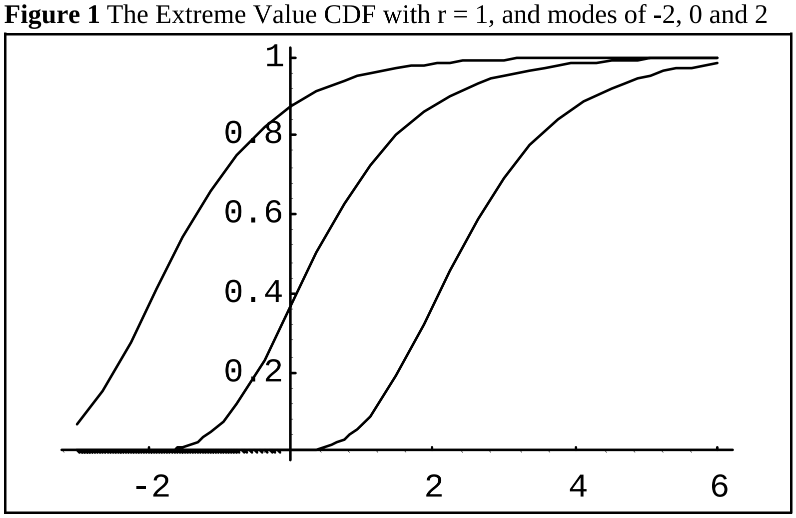
random elements for alternatives that are not in group *n*.[14] If, in addition, the amount the unobserved attribute **varies** within a group varies by group, alternatives in some groups will be more correlated with each other than the alternatives in other groups are correlated with each other. In these two cases, it is inappropriate to assume that the random terms for all C alternatives are independently drawn from the same univariate distribution. The NL CDF, Equation (8), allows the random terms to be correlated by groups and for the degree of correlation to vary by group.

For example, in our **model of participation and site choice**, one might expect the random terms in the conditional indirect utility function for the fishing sites ($<\epsilon_{1i}>$) to be more correlated with one another than they are with the random term for staying at home (ϵ_{21}) or the random term for bowling (ϵ_{22}). This would happen if there are important unobserved attributes of the alternatives that vary more, or less, across the fishing sites than they vary across the fishing sites and the other two alternatives. For example, the attribute fish stock varies across fishing sites but is always zero for staying at home and bowling, so omitting it would cause the ($<\epsilon_{1i}>$) to be more correlated with one another than they are with the (ϵ_{21}) or (ϵ_{22}).

Consider our **model of site choice with saltwater sites, lakes, and rivers.** Assume that aquatic vegetation varies significantly across rivers but not as much across lakes sites or saltwater sites. If this attribute of sites is unobserved, the random terms for the river alternatives ($<\epsilon_{3i}>$) will be more correlated with each other than with the random terms for lakes and saltwater sites.

Alternatively, if the **variation** in the unobserved attributes is not systematic by group type, it is reasonable to assume that each element of $<\epsilon_{mj}>$ is drawn from the same univariate distribution. This is what is assumed by the ML

Figure 1 The Extreme Value CDF with r = 1, and modes of -2, 0 and 2



[14]For example, if size is an important unobserved attribute of the alternatives, and size varies less within groups than across groups, the random terms for the alternatives that belong to group *n* will be more correlated with each other than they are with the random terms that are not in group *n*.

model. It assumes that each element of <ϵ_{mj}> is independently drawn from a univariate extreme value distribution

(25) $$F(\epsilon_{ni}) = \exp[-a_n e^{-r(\epsilon_{ni})}] = \exp[-e^{-r(\epsilon_{ni} - \alpha_n)}].$$

Recollect that $\alpha_m = \ln(a_m)$. The probability of choosing alternative *ni*, Equation (16) simplifies to

(26)[15] $$Prob(ni) = \frac{a_n e^{rV_{ni}}}{\sum_{m=1}^{M} \sum_{j=1}^{J_M} a_n e^{rV_{mj}}} = \frac{e^{r(\alpha_n + V_{ni})}}{\sum_{m=1}^{M} \sum_{j=1}^{J_M} e^{r(\alpha_n + V_{mj})}}.$$

[15]Looking back, Equation (19) looks like Equation (26) with $a_n = 1 \ \forall \ M$ and $s = r$. However, s in Equation (19) is not *r* in Equation (26). The s_n in Equation (19) can be viewed as s_n/r with *r* set equal to 1.

Diversion 3: Properties of the extreme value distribution [Equation (25)]. Precisely speaking, Equation (25) is a type 1 extreme value distribution with mode = α_n, mean = $\alpha_n + \gamma/r$, median = $\alpha_n + .6350\gamma/r$, and $\sigma^2 = \pi^2/6r^2$, where γ is Euler's constant (≈ .57721). Note that σ^2 $.57721 \pm 2.56$ increases as r decreases; e.g., if r = 1 and α_n = 0, 98% of the ϵ_{ni} are in the interval , whereas if r = .1 and α_n = 0, 98% of the ϵ_{ni} are in the interval 5.7721 ± 128 .

The density function is $f(\epsilon_{ni}) = re^{-r(\epsilon_{ni}-\alpha_n)}\exp[-e^{-r(\epsilon_{ni}-\alpha_n)}]$.

It is called an extreme value distribution because it can be derived as a distribution of extreme (largest or smallest) values. The distribution of maximum I.Q. is a distribution of extreme values, so is the distribution of maximum utility across discrete alternatives. However, the fact that the extreme value distribution can be interpreted as a distribution of extreme values has nothing to do with its choice as the appropriate distribution for the random component of utility for each alternative in a discrete choice model. Rather, its appeal is that its adoption leads to a closed form solution for *Prob(ni)*.

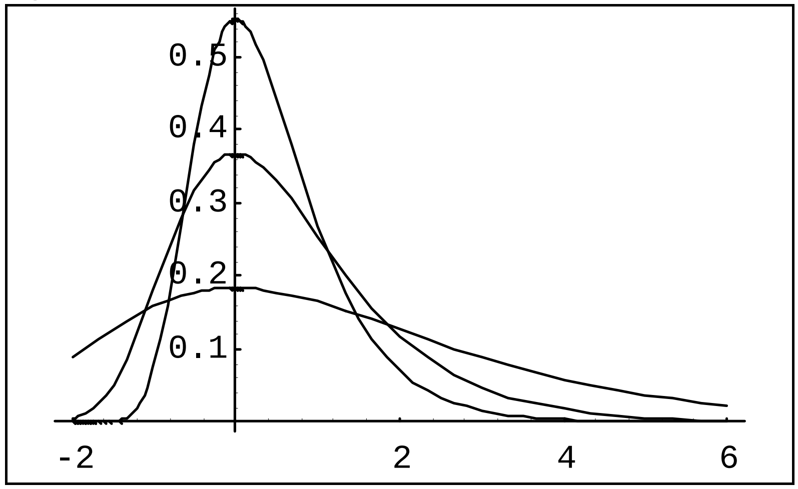
Note that with the ML model r can vary across individuals either as a function of the individual's characteristics or by allowing each individual to have a separate r. When there is variation in r, it is possible to estimate them. Estimation of a separate r for each individual requires multiple observations per individual. There are numerous reasons why one might want to allow r to vary: ability to make choices varies, the incentive to make choices can vary (e.g., with hypothetical versus actual choices), and the model includes all of the important explanatory variables for only some individuals. Morey and Rossmann (1997) estimate and discuss individual specific r's for two ML models. The estimated r will be relatively large (σ^2 small) for those individuals whose choices are being correctly predicted by $<V_{mj}>$ and small for those individuals whose choice cannot be explained by the $<V_{mj}>$. Note that $(Prob(ni) \rightarrow 1/C \;\forall\; ni \;\; as \;\; r \rightarrow 0)$.

Incorporating r's that vary by individual: If $U_{hij} = V_{hij} + \epsilon_{hij}$ where each ϵ_{hij} for individual h is an independent draw from $f(\epsilon) = \exp(-e^{-r_h(\epsilon)})$, then $\hat{U}_{hij} = r_h V_{hij} + \hat{\epsilon}_{hij}$ where each $\hat{\epsilon}_{hij}$ for individual h is an independent draw from $f(\epsilon) = \exp(-e^{-(\epsilon)})$, and the probabilities are Equation [26] with r replaced by r_h.

One might also allow r to vary across choice occasions as a function of the complexity of the problem [Swait and Adamowicz (1996)].

Note that with ML models it is typical, but unnecessary, to assume $a_m = 1 \;(\alpha_m = 0) \;\forall m$; that is, groups (represented by -group-specific constants) are consistent with the ML model. If the $<V_{mj}>$ are linear in their parameters, one cannot identify r separately from those parameters, so, without loss of generality, r is set to 1.[16] With $r=1$, Equation (26) can be derived from Equation (16) by restrictively that $s_m = 1 \ \forall m$.

Figure 2 The Extreme Value PDF with mode = 0, and r = .5, 1.0 and 1.5



As is well known, the ML model imposes the IIA assumption which says that the ratio of any two probabilities is independent of any change in any third alternative; that is[17],

$$(27)\quad \frac{Prob(ni)}{Prob(lk)} = \frac{e^{V_{ni}}}{e^{V_{lk}}} .$$

This restriction is correct if the variation in unobserved attributes is not systematic by group type, but inappropriate if it is.

If an important unobserved attribute has the same magnitude for all the alternatives in a group but differs across groups, this will affect the $<a_{mj}>$, but not the $<s_{mj}>$. Such attributes cause alternatives within a group to be more similar to one another than they are to alternatives in different groups, but does not influence the correlations of the $<s_{mj}>$. Consider the following example: size varies across alternatives, but

---

[16]E.g. if $V_{mj} = \mu(I - p_{mj}) \quad \forall mj$, where I is income and p_{mj} is the cost of the alternative, then $Prob(ni) = \frac{e^{(r\mu)(I - p_{ni})}}{\sum_{m=1}^{M} \sum_{j=1}^{J_M} e^{(r\mu)(I - p_{mj})}}$ and one cannot identify both r and μ

[17]In what follow r is set equal to 1.

all alternatives of the same type are the same size. The omission of size will cause the $<a_m>$ to vary in magnitude, but will not cause the elements of $<s_m>$ to differ from 1.

The a_m and s_m add systematic variation across groups that is in addition to the systematic variation in terms of the V_{mj}; that is, they allow the groups to differ in systematic ways in addition to the differences that can be attributed to variations in the observed attributes that appear as independent variables in the V_{mj}. a_n reflects the relative attractiveness of alternatives of type n. Ceteris paribus, a_n will be large, in a relative sense, if alternatives of type n have more of an important, but unobserved, attribute.

Note that allowing a_m to vary, while holding $s_m = 1 \; \forall \; M$, is not sufficient to weaken the IIA assumption. This can be seen by considering a case where $s_m = 1 \; \forall \; M$ but a_m varies. In this case,

$$(28) \quad Prob(ni) = \frac{a_n e^{V_{ni}}}{\sum_{m=1}^{M} a_m \sum_{j=1}^{J_m} e^{V_{mj}}} \quad \text{and} \quad \frac{Prob(ni)}{Prob(lk)} = \frac{a_n \; e^{V_{ni}}}{a_l \; e^{V_{lk}}} .$$

Thus, the IIA assumption still holds for all pairs of alternatives given that a_n and a_l are treated as parameters. IIA remains because the $<a_m>$ do not cause the elements of $<\epsilon_{mj}>$ to be correlated; that is, the model is still ML.

The s_n parameter, not a_n, is what is picking up part of the common (correlated) component in the random terms for all the alternatives of type n. The $<s_m>$ determine the extent to which the IIA assumption is imposed. Note that it is possible to let $<s_m>$ vary across individuals as, for example, a function of individual characteristics. This would allow the strength of the nest to vary across individuals but not the groupings.[18] As noted above, $s_m = 1 \; \forall$ M imposes IIA across all pairs of alternatives. Alternatively, if $s_m \neq 1 \; \forall \; M$, the IIA assumption will not be imposed across all pairs of alternatives, just some pairs.

---

[18]If the estimated (or specified) $<s_m>$ were all exactly one for some individuals, for this subset of individuals the nest would collapse and the model would be ML. A model can be NL for some individuals and ML for others.

Consider the two-level NL model with $a_m = a$ and $s_m = s \neq 1 \; \forall$ M. In this case *Prob(ni)*, Equation (16), simplifies to

$$(29) \quad Prob(ni) = \frac{e^{sV_{ni}} [\sum_{j=1}^{J_n} e^{sV_{nj}}]^{(1/s)-1}}{\sum_{m=1}^{M} [\sum_{j=1}^{J_m} e^{sV_{mj}}]^{1/s}} .$$

If $s_m = s \neq 1 \; \forall m$, the random terms in each group are more correlated with each other than they are with the random terms in other groups, but the degree to which they are more correlated with their fellow group members is constant across groups. In NL models of recreational demand, it is common to assume $s_m = s \neq 1 \; \forall m$.

When $a_m = a$ and $s_m = s \neq 1 \; \forall M$,

$$(30) \quad \frac{Prob(ni)}{Prob(lk)} = \frac{e^{sV_{ni}}[\sum_{j=1}^{J_n} e^{sV_{nj}}]^{(1/s)-1}}{e^{sV_{lk}}[\sum_{j=1}^{J_l} e^{sV_{lj}}]^{(1/s)-1}} .$$

Examining Equation (30), one sees that:

1. IIA still holds for any pair of alternatives within the same group ($n = l$). If $n=l$, Equation (30) simplifies to $\frac{Prob(ni)}{Prob(nk)} = \frac{e^{s_n V_{ni}}}{e^{s_n V_{nk}}}$.

2. IIA still holds for all pairs of alternatives that are in different groups ($n \neq l$) if the alternative changed is not in the same group as either alternative in the pair. That is, Equation (30) is not a function of changes in alternatives that are in other groups.

3. But, IIA does not hold for pairs of alternatives that are in different groups ($n \neq l$) if the alternative changed is in the same group as one of the alternatives in the pair. That is, Equation

For Example: Consider **the model of site choice with saltwater sites, lakes and rivers**, where *M* has three elements: 1= saltwater fishing, 2=lake fishing; and 3=river fishing; J_1=3, J_2=4 and J_3=5.

$$F(<\epsilon_{mj}>) = \exp\{-(a_1[\exp(-s_1\epsilon_{11})+...+\exp(-s_1\epsilon_{13})]^{(1/s1)} + a_2[\exp(-s_2\epsilon_{21})+... +\exp(-s_2\epsilon_{24})]^{(1/s2)} + a_3[\exp(-s_3\epsilon_{31})+...+\exp(-s_3\epsilon_{35})]^{(1/s3)})\}$$

The ratio of the probabilities for any two sites of the same type will not be influenced by a change in the attributes of any other site or sites. For example, the ratio of probabilities for two of the rivers will not change if the attributes of any of the saltwater sites, lakes, or other rivers change.

The ratio of the probabilities for any two sites that are different water types will not be influenced by a change in the attributes of a site (or sites) that is not of one of those water types. For example, the ratio of probabilities for a lake site and a river site will not change if the attributes of one or more saltwater sites change.

The ratio of the probabilities for any two sites that are different water types will be influenced by a change in the attributes of a site (or sites) that is one of these water types. For example, the ratio of probabilities for a lake site and a river site will change if the attributes of some other lake or river site change.

(30) is a function of the attributes of alternatives n and l, so a change in any alternative in either group n or l will affect the ratio.

For example, if alternative 11 is altered, it will not change $\frac{Prob(12)}{Prob(13)}$, $\frac{Prob(21)}{Prob(22)}$ or $\frac{Prob(31)}{Prob(42)}$, but it will affect $\frac{Prob(12)}{Prob(22)}$. In summary, generalizing from ML to NL only partially relaxes the IIA assumption. This is an important but often overlooked point. Generalizing Equation (29) by allowing both a_m and s_m to vary buys no more in terms of the IIA; IIA is still imposed to the same degree.

II. Estimation

Although this monograph is not about estimation per se, a few comments about estimation are in order. The log of the likelihood function [examples are Equation (21), (23) and (24)] can be maximized in one step by using a numerical algorithm to find the vector of parameters {$a_1, a_2,...,a_M; s_1, s_2,...,s_M$; and the parameters in the $<V_{mj}>$ functions} that maximize it. This approach is deemed Full Information Maximum Likelihood, FIML.[19] Alternatively, one can adopt a two-stage sequential estimation (SE). The parameters in the V_{mj} for group m can be divided into two categories for purposes of estimation, those that just influence the allocation between alternatives in group m (that is, those that appear in the conditional probabilities, Equation (19), for group m), and all other parameters. In the first step of sequential estimation, one sets $s = 1$ and for each group one separately estimates just those parameters that determine the allocation amongst the alternatives in that group. This is done by maximizing the log of the likelihood function for the choice of each j in the group conditional on the choice of n.[20] In this first stage, the model for each group is a ML model with J_n alternatives. In the second stage of a sequential estimation, one estimates the $<a_m>$, $<s_m>$, and other parameters, given the parameter estimates from the first stage.

Although this two- step estimation procedure is tempting, and it is the easier approach given existing computer hardware and software, I recommend against it. It has been known for a long time that the sequential technique leads to parameter estimates that are not asymptotically efficient, and, without a difficult correction, standard-error estimates that are inconsistent (Amemiya, 1978).[21] Although there is no result that states that FIML estimates always have better statistical properties in finite samples, FIML is

---

[19] Note that identification requires that one of the a_m is set to some positive constant; e.g. 1. If one's intent is to estimate a basic two-level nested model ($a_m = a$ and $s_m = s$ $\forall$), the a cancels out and the parameter vector is just {s; and the parameters in the $<V_{mj}>$}.

[20] Note that the data used to estimated the parameters in each nest are typically only a subset of the full data set because they include only those observations that involve a choice in that nest.

[21] Formula for correcting the standard errors can be found in Amemiya (1978) and McFadden(1981).

both consistent and asymptotically efficient, so it dominates in large samples, and FIML and sequential estimation often produce very different parameter estimates.[22] Monte Carlo studies also indicate that FIML estimates are more efficient than sequential estimation; see, for example, Brownstone and Small (1989).[23] Software programs (e.g., *Gauss* and other such programs) that directly maximize the log of the likelihood function for the full model, Equation (21), (23) or (24), are now widely available for both PCs and mainframes. There are numerous examples of FIML NL estimation both in other fields and in recreational demand. Two recreational examples are Morey *et al.* (1993 and 97a).

An important issue with FIML estimation of a NL model is the starting values for the parameters. For NL models, the FIML log-likelihood function is not guaranteed to be globally concave, so depending on the starting values for the parameters, one can possibly achieve a local rather than a global maximum.[24] This suggests that one estimate the model with a number of different starting values to increase the probability that the maximum found is the global maximum. Two set of possible starting values are the sequential estimates and the estimates derived by assuming the model is ML ($s_m = 1 \ \forall m$). The sequential estimates, although not asymptotically efficient, are consistent, so they are one set of reasonable starting values, particularly if the sample size is large.[25]

FIML estimation can often be quite difficult; the ability to find the maximum can be quite sensitive to starting values, the search algorithm, and how the parameters in the model are scaled. The starting values can affect the outcome in two ways: if there are multiple, local maxima, the one found often depends on

---

[22]See, for example, Cameron (1982 and 85), Hensher (1986), Brownstone and Small (1989), and Kling and Thompson (1996).

[23]In addition, simulations show that even when data have been generated by a NL model, SE estimates do not always exist.

[24]In contrast, the likelihood function for ML models is globally concave, so with ML models one needs to worry less about starting values.

[25]Hensher (1986) discourages the use of the sequential estimates as starting values for FIML estimation. In my opinion, they should not be used as the only starting values.

where the search begins, and starting values can cause search algorithms to search in an area of parameter space where small changes in parameter values can generate values of machine zero for components of the likelihood function, in which case the program crashes.[26] Maximum likelihood is invariant to rescaling parameters in the model (e.g., changing the units in which explanatory variables are denoted), but the search algorithms are not. Scaling determines the relative rates at which the parameters change from iteration to iteration. Search algorithms generally converge more quickly when the explanatory variables are scaled such that their parameters are of the same order of magnitude. A characteristic of NL models is that the log likelihood is highly non-quadratic in the $<s_m>$ parameters. This makes the $<s_m>$ parameters difficult to estimate, and implies that one should avoid search algorithms that rely on quadratic approximations (e.g. BHHH). In addition, it is probably wise to scale the $<s_m>$ parameters such that they vary more slowly than the other parameters.

Estimation often leads to an estimate(s) for s (s_m) that is greater than zero but less than one. This result indicates that the estimated NL model is not globally consistent with stochastic utility maximization (⇔ not globally consistent with an underlying density

Diversion 4: Conditions under which the estimated model is locally consistent with stochastic utility maximization at $<V_{mj}>$.

A necessary, but generally not a sufficient condition for local consistency is that $Prob(k) \geq 1 - s_k \ \forall\ k$, where *Prob(k)* is the probability of choosing an alternative in group *k*. Note that this condition will always be fulfilled if $(s_m \geq 1 \ \forall\ m)$, but if $(0 < s_k < 1)$ the probability of the condition being violated increases as *Prob(k)* decreases and as s_k decreases.

If there are J_k alternatives in group *k*, then there are $(J_k - 1)$ conditions for that group, including $Prob(k) \geq 1 - s_k$. For details see Herriges and Kling (1996), and remember that they express the conditions in terms of θ_k *,where* $\theta_k = 1/s_k$.

[26]One can increase the accuracy of the likelihood function and reduce the probability of encountering machine infinity by coding expressions of the form $ln(e^a + e^b + e^c)$ as $a + \ln(1 + e^{b-a} + e^{c-a})$.

function).[27] For details see *Diversions 1* and *2*. For cases where the s_m are all positive but not all one or greater, Börsh-Supan (1990) and Herriges and Kling (1996) have derived the conditions under which the estimated NL model is consistent with stochastic utility maximization at a specified $<V_{mj}>$.[28] If these local conditions hold at every $<V_{mj}>$ in the data set and at every $<V_{mj}>$ that is being evaluated for policy purposes, the estimated model is locally consistent with stochastic utility maximization. The basics of these conditions are presented in *Diversion 4,* but an example from Kling and Herriges (1995) suggests that if some of the estimated s_m are less than one by a significant amount, the estimated model will not be locally consistent with utility maximization.

When the estimated model is neither globally or locally consistent with stochastic utility maximization, it is an indication that the NL framework is not appropriate, or the assumed nesting structure is incorrect for the population, or the assumed nesting structure is incorrect for some subset of the population, or the $<V_{mj}>$ are miss-specified. One can impose global consistency by replacing, in estimation, s_m with $s_m = 1 + e^{\rho_m}$. Not surprisingly, imposing consistency can have a significant effect on the parameter estimates. For more specifics on the implications of different nesting structures, see Kling and Thompson (1996) and Hauber and Parsons (1996).

One must also be cognizant of choice-based samples; that is, samples where individuals are selected into the sample as a function of the alternatives(s) they choose. For example, a sample of anglers recruited at fishing sites is choice-based, and the sample anglers will take more trips, on average, than anglers randomly chosen from the population of anglers. Choice-based samples are typically easier to collect than a random sample of the population.

---

[27] To complicate things more, note that since the estimated s_m are random variables, one should be concerned with whether one can reject the joint null hypothesis that the $<s_{mj}>$ are each 1 or greater. Note that this complication also holds when considering local consistency. For more details see Kling and Herriges (1995).

[28] Herriges and Kling (1996) correct a misstatement in Börsh-Supan (1990) and provide second order derivative conditions and some bounds on the consistency region.

Without weights in the likelihood function to correct for the non-representativeness of the sample in terms of choices made, parameter estimates will be inconsistent. Simply put, choices need to be weighted such that the contribution to the likelihood function is reduced for choices that are over-represented in the sample and increased for choices that under-represented in the sample. With correct weights choices in the weighted sample have the same proportions as in the population. For details see Manski and Lerman (1977). Morey *et. al.* (1995 and 1997a) use weights to correct for the fact that the anglers in their choice-based sample take, on average, more trips than the average angler in their target population. Creation of weights requires independent information on how alternative choices are distributed in the population, and this information is often not available.

One must also be cautious when using the estimated parameters to draw inferences from a sample to the population. If the sample is not representative of the population in terms of the characteristics of its members, sample averages will not reflect population averages, even when one has correctly estimated the parameters.[29] To estimate behavioral changes or consumer's surplus estimates for the target population, one need to know the joint distribution in the population of the individual characteristics in the model. If this distribution is known, often it is not, one can determine the probability that an individual with certain characteristics appears in the population and compare this to the proportion of individuals in the

> **For example**, **if the intent is to estimate the average consumer's surplus for a population of anglers**, if gender is the only individual characteristics in the model, 60% of anglers in the population are male but 80% of the anglers in the sample are male, then to obtain an estimate of the average consumer's surplus for the population, before averaging, the female consumer's surplus estimates in the model should be multiplied by two and the male estimates by 3/4.
>
> For example, Morey *et al.*(1997b), when estimating average consumer's surplus for the population, weight to correct for the fact that low income households, blacks, and certain age groups are under-represented in the sample.

[29]It is important to distinguish between samples that are not representative in terms of the choices made verus samples that are not representative in terms of important individual characteristics. Samples can be either, neither or both. Nonrepresentative in terms of choice made causes, without weighting, parameter estimates to be inconsistent, nonrepresentative in term of individual characteristics does not.

sample with these characteristics. This information can then be used to weight the estimate for each individual in the sample to obtain a weighted average that reflects the population average.

In conclusion, one needs to consider why there should be group constants or a nested model. If all of the important explanatory variables are included in the model, there is no need for group or alternative specific constants, and they should not be included. Including group or alternative specific constants will always improve the fit but can mask the true effects of the included explanatory variables. The more numerous the added constants, the less there is left to be explained by the explanatory variables. The intent of modeling and estimation is to explain observed behavior and how behavior and utility will differ if the values of the explanatory variables change, not to solely predict the sample, so constants should be used with caution.

One should also ask why the random components are likely to be correlated by group. If the answer is an omitted variable that varies more (or less) within a group than across groups, it would be better to try and collect that data and include the variable in the model than to nest. Nesting can also mask the effect of explanatory variables.

IV. Derivation of Expected Maximum Utility for the Two-level NL Model

The intent of this section is to use the basics of probability theory to derive expected maximum utility from Equations (1) and (8). Expected maximum utility often plays an important role in the derivation of compensating variation and equivalent variation. Before proceeding with the derivation of expected maximum utility from the NL model, it is important to point out that the expected maximum utility derived in this section is the expected maximum utility *per choice occasion*, not for the year or fishing season. One must remain cognizant of this if one's intent is to derive the per year compensating variation associated with a change in the attributes of a site or sites.

Let $U = max(<U_{mj}>) \equiv max(<V_{mj} + \epsilon_{mj}>)$ denote the largest element in the vector $<V_{mj} + \epsilon_{mj}>$. Therefore, given Equation (1), expected maximum utility, $E(U)$, is

(31)

$$E(U) \;=\; \int_{\epsilon_{11}=-\infty}^{+\infty} \int_{\epsilon_{12}=-\infty}^{+\infty} ... \int_{\epsilon_{mj}=-\infty}^{+\infty} ... \int_{\epsilon_{MJ}=-\infty}^{+\infty} \max(V_{11}+\epsilon_{11}, V_{12}+\epsilon_{12}, ..., V_{mj}+\epsilon_{mj}, ..., V_{MJ}+\epsilon_{MJ})$$

$$f(\epsilon_{11},\epsilon_{12},...,\epsilon_{mj},...,\epsilon_{MJ}) d\epsilon_{MJ} ... d\epsilon_{mj} ... d\epsilon_{12} d\epsilon_{11} \;.$$

Equation (31) is the equation for the expected value of the function $max(<V_{mj} + \epsilon_{mj}>)$. Recall that the individual knows his or her maximum utility; $E(U)$ is our expectation of it. Equation (31) can be written more simply by dividing the density into *C* regions such that in region *ni* alternative *ni* is chosen (i.e., in region *ni*, alternative *ni* has maximum utility). Dividing into these *ni* regions, one obtains

$$(32) \quad E(U) \;=\; \sum_{n=1}^{M} \sum_{i=1}^{J_n} \int_{\epsilon_{ni}=-\infty}^{+\infty} (V_{ni} \;+\; \epsilon_{ni}) \;\; F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>) d\epsilon_{ni} \;,$$

where, as noted in Equation (7) $\int_{\epsilon_{ni}=-\infty}^{+\infty} F_{ni}(<V_{ni}+\epsilon_{ni}-V_{mj}>) d\epsilon_{ni} \;=\; Prob(ni)$. Equation (32) identifies expected maximum utility for **any** discrete choice model that is consistent with Equation (1). In this sense Equation (32) is quite general, and one could, in theory, plug any specific CDF, $F(<\epsilon_{mj}>)$, into Equation (32) to derive the expected maximum utility associated with that CDF. A critical issue, as with the derivation of the *Prob(ni)*, is when Equation (32) will have a closed-form solution. It has a closed-form solution if one assumes the generalized extreme value distribution denoted in Equation (8), as is now demonstrated.

To obtain expected maximum utility for the two-level NL model, substitute Equation (10) into Equation (32) to obtain

$$(33) \quad E(U) \;=\; \sum_{n=1}^{M} \sum_{i=1}^{J_n} \int_{\epsilon_{ni}=-\infty}^{+\infty} (V_{ni} \;+\; \epsilon_{ni}) \;\; \exp\{-\sum_{m=1}^{M} a_m [\sum_{j=1}^{J_m} e^{-s_m(V_{ni}+\epsilon_{ni}-V_{mj})}]^{(1/s_m)}\}$$

$$a_n [\sum_{j=1}^{J_n} e^{-s_n(V_{ni}-V_{nj}+\epsilon_{ni})}]^{(1/s_n)-1} e^{-s_n \epsilon_{ni}} \;\; d\epsilon_{ni} \;.$$

Simplify by making the change of variables $w = V_{ni} + \epsilon_{ni}$ ($\Rightarrow d\epsilon = dw$ because V_{ni} is a constant, and $\epsilon_{ni} = w - V_{ni}$) to obtain

$$(34)\quad E(U) = \sum_{n=1}^{M}\sum_{i=1}^{J_n}\int_{w=-\infty}^{+\infty} w\ \exp\{-\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{-s_m(w - V_{mj})}]^{(1/s_m)}\}$$

$$a_n[\sum_{j=1}^{J_n} e^{-s_n(w - V_{nj})}]^{(1/s_n)-1} e^{-s_n(w - V_{ni})}\ dw.$$

Note that the term in Equation (34), $\exp\{-\sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{-s_m(w - V_{mj})}]^{(1/s_m)}\}$

$$= \exp[-De^{-w}]$$

where

$$(35)\quad D = \sum_{m=1}^{M} a_m[\sum_{j=1}^{J_m} e^{s_m V_{mj}}]^{(1/s_m)},$$

and that the term in Equation (34),

$$[\sum_{j=1}^{J_n} e^{-s_n(w - V_{nj})}]^{(1/s_n)-1}$$

$$= e^{-w}\ e^{s_n w}\ [\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1}.$$

Substituting these two simplifications into Equation (34) and moving all of the terms that do not contain w, with the exception of D, to the left of the integral sign, one obtains

$$(36)\quad E(U) = \{\sum_{n=1}^{M}\sum_{i=1}^{J_n} a_n e^{s_n V_{ni}}[\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1}\}$$

$$\int_{w=-\infty}^{+\infty} we^{-w}\exp\{-De^{-w}\}dw$$

Now examine the first term in Equation (36). It equals *D*. This follows from Euler's Theorem, as is now demonstrated. *D*, Equation (35), is homogenous of degree one in $< e^{V_{mj}} >$, and

$$(37)\quad \frac{\partial D}{\partial(e^{V_{ni}})} = e^{-V_{ni}} a_n e^{s_n V_{ni}} [\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1} \quad .$$

Therefore, by Euler's Theorem[30],

$$(38)\quad \sum_{n=1}^{M} \sum_{i=1}^{J_n} e^{V_{ni}} \frac{\partial D}{\partial(e^{V_{ni}})} = \sum_{n=1}^{M} \sum_{i=1}^{J_n} e^{V_{ni}} \{ e^{-V_{ni}} a_n e^{s_n V_{ni}} [\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1} \}$$

$$= \sum_{n=1}^{M} \sum_{i=1}^{J_n} a_n e^{s_n V_{ni}} [\sum_{j=1}^{J_n} e^{s_n V_{nj}}]^{(1/s_n)-1} = D \quad .$$

Since the first term in Equation (36) equals *D*, Equation (36) implies

$$(39)\quad E(U) = \int_{w=-\infty}^{+\infty} w D e^{-w} \exp\{-D e^{-w}\} dw \, .$$

This is where things get exciting. Note that the density function for an extreme value distribution with mode = *ln(D)* is[31]

$$(40)\quad f(w) = D e^{-w} \exp\{-D e^{-w}\} \, .$$

---

[30]Euler's Theorem states $D = \sum_{n=1}^{M} \sum_{i=1}^{J_n} \alpha_{ni} \frac{\partial D}{\partial \alpha_{ni}}$ if *D* is homogenous of degree one in the $< \alpha_{mj} >$. In our case, $\alpha_{ni} = e^{V_{ni}}$.

[31]For more details on the extreme value distribution see Equation (25), *Diversion* 3 and Johnson *et al.* (1994)).

Therefore, since $E(w) = \int_{w=-\infty}^{+\infty} wf(w)dw$, Equation (39) is the expected value (mean) of an extreme value distribution with a scale parameter of 1. It is well known that this expected value is ln*D* + .5772..., where .57721... is Euler's Constant. Therefore, expected maximum utility is

$$(41) \quad E(U) = \ln D + .57721.... \ .$$

E(U) can be interpreted as the expected maximum utility for a representative individual. Equation (41) is used in Section VI to derive a compensating variation and an equivalent variation from the NL model. Remember that *E(U)* is the expected maximum utility per choice occasion and that there will be a different *E(U)* for each type of individual.

> **For example:** In the **model of participation and site choice**, the season is divided into a finite number of periods (choice occasions) so *E(U)* is expected maximum utility per choice occasion. Expected maximum utility for the season is the sum of the per-period expected utilities. If attributes vary from period to period, expected utility will vary across the periods.
>
> In the **model of site choice with saltwater sites, lakes and rivers**, each trip is a choice occasion so *E(U)* is the expected maximum utility **per-trip**. Per-trip expected maximum utility does not easily translate into seasonal expected maximum utility.
>
> As we will see in the derivation of compensating variation, the distinction between per-period and per-trip expected maximum utility is quite important.

As an aside, note that it can be shown that

$$(42) \quad \frac{\partial E(U)}{\partial V_{ni}} = Prob(ni).$$

Said loosely, Equation (42) is the discrete-choice random-utility analog of Roy's Identity where V_{ni} is interpreted as the negative of the *normalized price* of alternative *ni*.

V. The Functional Forms of the Conditional Indirect Utility Functions, $<V_{mj}>$, and Budget Exhaustion

Let *I* denote the budget for the choice occasion, p_{mj} the price (cost) of alternative *mj*, β^{mj} the vector of other observed attributes of alternative *mj*, and *G* the vector of characteristics of the individual, excluding budget, that influence the choice of alternative. The choice set consists of *C* mutually exclusive alternatives and a numeraire composite good. During each choice occasion, the individual is constrained to consume one, and only one, of the C alternatives, and then spends the rest of his or her budget on the numeraire good. For example, if alternative *ni* is chosen, $(I - p_{ni})$ is spent on the numeraire good and budget exhaustion requires that V_{mj} is restricted to functions of the form[32]

(43) $$V_{mj} = V_{mj}((I - p_{mj}), \beta^{mj}, G) \qquad \forall\ mj.$$

Denote $\partial V_{mj}/\partial(I - p_{mj}) = \mu((I - p_{mj}), \beta^{mj}, G)$, where $\mu(.)$ is the marginal utility of money, that is, the utility that the individual gets from the marginal unit of the numeraire. Regularity requires that $\mu(.)$ is nonnegative.

Consider first cases where $\mu(.)$ is not a function of $(I - p_{mj})$ or β^{mj}. Such models will be referred to as models with *zero income effects*, that is, models where, for a given individual, the marginal utility of money is a constant, independent of which alternative is chosen. For example,

(43a) $$V_{mj} = \mu(G)(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \qquad \forall\ mj.$$

Note that Equation (43a) allows the marginal utility of money to be different for different individuals; e.g., it could vary as a function of age, race, gender, or fishing ability.[33] Models consistent with Equation (43a)

---

[32]More generally, $V_{mj} = V_{mj}((I - p_{mj}), p_{num}, \beta^{mj}, G)$ where p_{num} is the price of the numeraire. Since this function is an indirect utility function, it must possess all the required properties of such functions. These properties are continuity, homogeneity of degree zero in (I, p_{num}, p_{mj}), nondecreasing in $(I - p_{mj})$, and nonincreasing and quasiconvex in p_{num}. Assuming all individuals in the population face the same price for the numeraire, p_{num}, is set equal to one without loss of generality, and all $V_{mj} = V_{mj}((I - p_{mj}), \beta^{mj}, G)$ are consistent with $V_{mj} = V_{mj}((I - p_{mj}), p_{num}, \beta^{mj}, G)$ being homogeneous of degree zero in (I, p_{num}, p_{mj}), and nonincreasing and quasiconvex in p_{num}. Regularity then requires only that $V_{mj} = V_{mj}((I - p_{mj}), \beta^{mj}, G)$ be continuous in $((I - p_{mj}, \beta^{mj})$ and nondecreasing in $(I - p_{mj})$.

[33]Some of these individual characteristics might be correlated with the budget. Buchanan *et al.* (1997), in a model to explain the choice of mountain bike sites, find μ to be a function of the individual's characteristics, so does Morey et al. (1997b), in a study valuing the preservation of monuments, where

(continued...)

are called *zero income effects* models because the budget, *I*, does not influence the choice probabilities. Choosing a model with *zero income effects* means that data on the budget per choice occasion need not be collected, but also means that compensating variations are constrained to not vary by income level.

The reason that the budget drops out of the choice probabilities is that any attribute (budget, price or other attribute) that adds the same constant to each of the *C* conditional indirect utility functions will not influence the choice of alternative [$\mu(G)I$ is one such constant].

Looking ahead, when there are *zero income effects*, the compensating and equivalent variations for any proposed change are equal and easy to estimate. Because of this, in the applied literature it is very common to assume that

(43a-1) $$V_{mj} = \mu_0(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \quad \forall \; mj,$$

where μ_0 is the same constant for all individuals. However, the more general Equation (43a) is often preferable because it allows for more variation in choices, and allows willingness to pay to vary across individuals as a function of their characteristics.

Things get more complicated when there are income effects. Models where $\mu(.)$ is a function $(I - p_{mj})$ or β^{mj} are deemed models with *income effects.* These are models where the marginal utility of money varies as *I* varies or because the utility from the numeraire depends on which of the discrete alternatives is chosen. I will consider just a few simple examples. Consider first a model where the marginal utility of money depends on which alternative is chosen but not on the budget.[34]

(43b) $$V_{mj} = \mu_{mj}(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \quad \forall \; mj.$$

---

[33](...continued)
gender is a significant determinant of μ.

[34]Note that Equation (43b) is a special case of $V_{mj} = \mu(\beta^{mj}, G)(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \quad \forall \; mj$.

A rationale for Equation (43b) is the marginal enjoyment the individual gets from units of the numeraire, although independent of the number of units consumed, is dependent on which alternative was chosen. This would result if some of the alternatives were more or less complementary with the numeraire.

> **Diversion 5: Consider the system of conditional indirect utility functions**
> (43b-1) $V_{mj} = \mu_m(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \quad \forall\ mj.$
> The budget, I, will influence whether the alternative chosen is of type m, but will not influence which alternative of type m will be chosen.
>
> **More generally**, consider some attribute of the individual, g_k that enters all of the conditional indirect utility functions as "$+\ \alpha g_k$". It will not influence the probabilities. Alternatively, if g_k interacts with one or more of the site characteristics, it will affect choice.

Now consider models where the marginal utility an individual gets from consuming the numeraire does not depend on which alternative was chosen but depends on the amount of the numeraire consumed. For example,

(43c) $$V_{mj} = \mu(G, (I - p_{mj}))(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \qquad \forall\ mj.$$

Consider two cases of Equation (43c): models where $\partial\mu(.)/\partial(I - p_{mj}) \neq 0\ \forall\ (I - p_{mj})$ - denote these as models with *continuous income effects;* and models where $\mu(G, (I - p_{mj}))$ is a step function in $(I - p_{mj})$ - denote these as models with *step income effects.* The distinction is important for data collection, estimation, and the calculation of the compensating variation. An example of a *continuous income effects* model is

(43c-1) $$V_{mj} = \mu_0(I - p_{mj}) + \mu_1(I - p_{mj})^2 + h_{mj}(\beta^{mj}, G) \qquad \forall\ mj.$$

Estimation of continuous income effects models requires data on the budget per choice occasion.[35]

An example of a *step income effects* model is

---

[35]Models with *continuous income effects* include Morey, *et al.* (1993), Buchanan (1997) and Herriges and Kling (1997). Morey *et al.* (1993) include all alternatives in the choice set and assume the budget per choice-occasion is yearly income divided by the specified number of choice occasions. Buchanan *et al.* (1997) estimate a model where the choice set includes only mountain bike sites, and they assume the daily budget available for mountain biking is 1/30 of monthly discretionary income.

(43c-2) $V_{mj} = g(I - p_{mj}) + h_{mj}(\beta^{mj}, G) \quad \forall\ mj.$

where

$$g(I - p_{mj}) = \mu_0 (I - p_{mj}) \text{ if } (I - p_{mj}) \leq I^0,$$

$$g(I - p_{mj}) = \mu_0 (I^0) + \mu_1 (I - p_{mj} - I^0) \quad \text{if } I^1 \geq (I - p_{mj}) > I^0$$

$$g(I - p_{mj}) = \mu_0 (I^0) + \mu_1 (I^1 - I^0) + \mu_2 (I - p_{mj} - I^1) \quad \text{if } (I - p_{mj}) > I^1,$$

Denote this function, $g(I - p_{mj})$, a step function of $(I - p_{mj})$, where μ_0 is the marginal utility of the first I^0 units of the numeraire, μ_1 is the marginal utility of the next $(I^1 - I^0)$ units of the numeraire, etc. There is no presumption that the sequence μ_0, μ_1, μ_2 is either monotonically increasing or decreasing.

Assuming *step income effects* is a simple and often realistic way to incorporate income effects. For example, if one believes that willingness to pay for a policy scenario remains constant over broad income ranges but does depend on whether one is poor, a two-step function would be appropriate. In addition, incorporating income effects in this manner does not require detailed budget data, just the appropriate budget category.

For example, in the **model of site choice with saltwater sites, lakes and rivers**, expected catch rate is likely to be an important attribute of all of the sites. Assume there is some variation in catch but that all lake sites have the same catch rate and the coefficient on catch is the same constant in all of the conditional indirect utility functions for lake sites. In this case, catch rate will influence whether one chooses a lake site, but, conditional on a lake trip being chosen, will not influence which specific lake is visited.

Returning to the general issue of how site attributes affect site choice, any site characteristic that enters all the conditional indirect utility functions as a linear term with a common parameter, and that varies in magnitude across alternatives but does not vary in magnitude across the alternatives in *m*, will influence whether an alternative of type *m* is chosen, but not which alternative in *m*. For example, if size is an attribute, all alternatives of type *m* are size 10, and the coefficient on size is the same in all of the conditional indirect utility functions for alternatives of type *m*, conditional on choosing an alternative of type *m*, size will not influence which of the alternatives of type *m* you will choose.

VI. Compensating Variation and Equivalent Variation

Let $P \equiv \langle p_{mj} \rangle$, $\beta \equiv \langle \beta^{mj} \rangle$, and consider a change from the initial state $\{I^0, P^0, \beta^0\}$ to some proposed state $\{I^1, P^1, \beta^1\}$. For the choice occasion, maximum utility in the initial state is

(44) $$U^0 \equiv \max(\langle U^0_{mj} \rangle) = \max(\langle V^0_{mj} + \epsilon_{mj} \rangle) \equiv U(I^0, P^0, \beta^0, G, \langle \epsilon_{mj} \rangle),$$

and maximum utility in the proposed state is[36]

(45) $$U^1 \equiv \max(\langle U^1_{mj} \rangle) = \max(\langle V^1_{mj} + \epsilon_{mj} \rangle) \equiv U(I^1, P^1, \beta^1, G, \langle \epsilon_{mj} \rangle).$$

Denote maximum utility in the proposed state with compensation, *c*, subtracted from the budget

(46) $$U^1(c) \equiv \max(\langle U^1_{mj}(c) \rangle) = \max(\langle V^1_{mj}(c) + \epsilon_{mj} \rangle) \equiv U(I^1 - c, P^1, \beta^1, G, \langle \epsilon_{mj} \rangle),$$

where

(47) $$V^1_{mj}(c) \equiv V_{mj}((I^1 - p^1_{mj} - c), \beta^{mj1}, G).$$

Note that $U^1(c{=}0) = U^1$. For the choice occasion, the compensating variation, CV, that the individual associates with a change from $\{I^0, P^0, \beta^0\}$ to $\{I^1, P^1, \beta^1\}$ is that *c* which equates U^0 and $U^1(c)$; that is,

(48) $$U(I^0, P_0, \beta^0, G, \langle \epsilon_{mj} \rangle) \equiv U(I^1 - CV, P^1, \beta^1, G, \langle \epsilon_{mj} \rangle).$$

The CV is known to the individual but is a random variable from the researcher's perspective; to know CV one must know $\langle \epsilon_{mj} \rangle$. For an improvement, the CV is nonnegative and often zero. For a deterioration, the CV is nonpositive and often zero. For example, if the change is a deterioration in alternative *lk* and an individual was not choosing *lk*, the individual's CV is zero. For this deterioration, the CV is negative for only those individuals that were choosing *lk*.

---

[36]Note that $\langle \epsilon_{mj} \rangle$ is assumed not state specific; that is, $\langle \epsilon^0_{mj} \rangle = \langle \epsilon^1_{mj} \rangle$, a questionable but universal assumption.

The equivalent variation, EV, is defined in an analogous manner. Denote maximum utility in the initial state with compensation, *cc*, added to the budget:

(49)

$$U^0(cc) \equiv \max(<U^0_{mj}(cc)>) = \max(<V^0_{mj}(cc) + \epsilon_{mj}>) \equiv U(I^0 + cc, P^0, \beta^0, G, <\epsilon_{mj}>),$$

where

(50) $$V^0_{mj}(cc) \equiv V_{mj}((I^0 - p^0_{mj} + cc), \beta^{mj0}, G).$$

Note that $U^0(cc{=}0) = U^0$. For the choice occasion, the equivalent variation, EV, that the individual associates with a change from $\{I^0, P^0, \beta^0\}$ to $\{I^1, P^1, \beta^1\}$ is that *cc* which equates $U^0(cc)$ and U^1 ; that is,

(51) $$U(I^0 + EV, P_0, \beta^0, G, <\epsilon_{mj}>) = U(I^1, P_1, \beta^1, G, <\epsilon_{mj}>).$$

Like the CV, the EV is known to the individual but is a random variable from the researcher's perspective.

The CV and EV are always of the same sign. For an improvement, the CV is how much the individual would be willing to pay, per choice occasion, for the improvement, and the EV is how much the individual would have to be paid to voluntarily forego the improvement. For a deterioration, the CV is how much the individual would have to be paid to voluntarily accept the deterioration, and the EV is how much the individual would be willing to pay to stop the deterioration.

The CV will vary across individuals for two reasons: the $<V_{mj}>$ can vary across individuals, and for individuals with the same $<V_{mj}>$, $<\epsilon_{mj}>$ will vary across individuals. The CV for a given individual will also vary across choice occasions because the $<\epsilon_{mj}>$ vary across choice occasions. The $<V_{mj}>$ will vary across individuals for many reasons: differences in income, different costs for alternatives, and differences in other individual characteristics that affect the $<V_{mj}>$. Denote all individuals who have the same $<V_{mj}>$ as individuals of the same *type*.

Each individual has a specific CV, but from the researcher's perspective the CV for individuals of a given *type* is a random variable with some density function, $f(CV)$. Note that although the CV for a given change can be positive for some individuals and negative for others, it will either be nonnegative or nonpositive, but not both, for all individual of the same type. The only exception would be a zero CV for all individuals of the same type. $f(CV)$ will have a finite range, and one end of the range will be zero. For example, for an improvement in alternative *mj*, all other alternatives remain unchanged, the CV will be zero for all those individuals of this type that do not choose alternative *mj* in the new state, and the maximum CV will be the CV for those who chose alternative *mj* before it was improved.[37] More generally, denote the CV for an individual that chooses *mj* both before and after a change $CV(mj|mj)$. $CV(mj|mj)$ is the *c* that equates $(V^1_{mj}(c) + \epsilon_{mj})$ *and* $(V^0_{mj} + \epsilon_{mj})$, so it does not depend on ϵ_{mj} and can be easily calculated.[38]

The median CV for individuals of a given *type*, CV^{med}, can be zero. For example, if alternative *mj* is improved, all other alternatives unchanged, and the probability of choosing alternative *mj* is less than .5 in the improved state ($Prob^1(mj) < .5$), $CV^{med} = 0$ because the majority of individuals of this type are not choosing *mj* in the new state. In contrast, if alternative *mj* is improved, all other alternatives unchanged, and $Prob^0(mj) > .5$, $CV^{med} = CV(mj|mj)$. For a deterioration in alternative *mj*, all other alternatives unchanged, $CV^{med} = 0$ if $Prob^0(mj) < .5$, and $CV^{med} = CV(mj|mj)$ if $Prob^1(jm) > .5$. Approximation of CV^{med} for probabilities not in these ranges is discussed below.

Denote the expected value of the CV for individuals of a given type, E(CV). In general, it is not possible to directly calculate E(CV) from Equation (48); the primary exception is when there are *zero income effects* or *step income effects*.[39] However, it is possible to approximate E(CV). Note that the range

---

[37]If the individual chooses alternative *mj* and then alternative *mj* is improved with no changes in the other alternatives, the individual will also choose *mj* after the change because *mj* maximizes utility in both states.

[38]In contrast, consider the CV for an individual that initially choose alternative *ni* and then switches to alternative *mj*, $CV(ni|mj)$. It depends on both ϵ_{ni} and ϵ_{mj}.

[39]It is possible to create examples where it is possible to solve for E(CV) even when there are income effects, but the problem quickly becomes intractable and is intractable for most estimated models. Consider a simple example with only two alternatives, where alternative 1 is improved and alternative 2 remains the same. In this case, there are only three possible behavior types: those who choose 1 both before and after the change, those that initially chose 2 but switch to 1, and those that choose 2 in both states. No one will switch from 1 to 2. One can easily calculate $CV(1|1)$, and $CV(2|2) = 0$. One can
(continued...)

of the CV provides upper and lower bounds on E(CV), and, as the above example indicates, these can sometimes be calculated. Denote the CV for the *representative* individual of a given type CV^R, where CV^R is the monetary compensation (or payment) in the proposed state that would make the expected maximum utility in the proposed state equal to the expected maximum utility in the initial state; that is,

$$(52)\quad E(U^0) \equiv E[U(I^0,\ P^0,\ \beta^0,\ G,\ <\epsilon_{mj}>)] = E[(U(I^1 - CV^R,\ P^1,\ \beta^1,\ G,\ <\epsilon_{mj}>)],$$

where expected maximum utility, *E(U)*, is defined by Equations (41) and (35). *E(U)* is the utility of a representative individual in that it is maximum utility if $\epsilon_{mj} = 0\ \forall mj$. CV^R is the random-utility analog of the standard definition of the compensating variation in terms of a continuous utility function.[40] The distinctions between E(CV), CV^R, and CV^{med} were first considered by Hanemann (1985). A revised version of that paper appears in this book as chapter 3.

Define the EV^R, per choice occasion, as

$$(53)\quad E(U^1) \equiv E[U(I^1,\ P^1,\ \beta^1,\ G,\ <\epsilon_{mj}>)] = E[U(I^0 + EV^R,\ P^0,\ \beta^0,\ G,\ <\epsilon_{mj}>)],$$

that is, the monetary compensation (or payment) in the initial state that would make expected maximum utility in the initial state equal to expected maximum utility in the proposed state.

---

[39](...continued)
also calculate the proportion of individuals that will choose each behavior type. The proportion of individuals that choose 1 is $Prob^0(1)$, and the proportion that choose 2 in both states is $Prob^1(2)$, so the proportion that switches from 2 to 1 is $1 - Prob^0(1) - Prob^1(2)$. Therefore $E(CV) = Prob^0(1)CV(1|1) + \theta$, where θ is $f(\epsilon_1,\ \epsilon_2)CV(2|1)$ integrated over that part of the density function where individuals of this type switch from 2 to 1. Even in this simple case, this might be a difficult integral to solve, but there are examples where it can be solved.

[40]To avoid confusion, note that the term *representative consumer* has two possible meaning in the ML and NL literature: the above meaning, and as a *consumer* whose utility function over continuous amounts of each alternative represents the aggregate behavior of a group individuals who are individually constrained to choose only one alternatively, but collectively choose some of each alternative. For details on this latter meaning see, Anderson *et al.* (1987 and 1992), Verboven (1996), and Smith and Von Haefen (1997)

If one imposes *zero income effects* [Equation (43a)], Equation (45) can be solved for CV^R, and $CV^R = EV^{R.}$ Specifically,

$$(54)\quad CV^R = EV^R = (1/\mu)[E[U(I^0, P^1, \beta^1, G, <\epsilon_{mj}>)] - E[U(I^0, P^0, \beta^0, G, <\epsilon_{mj}>)]] + (I^1 - I^0)$$

$$= (1/\mu)[\ln D^1 - \ln D^0],$$

where D is defined in Equation (35), D^1 is D evaluated at $\{I^1, P^1, \beta^1\}$, and D^0 is D evaluated at $\{I^0, P^0, \beta^0\}$. Intuitively, in the case of *zero income effects*, CV^R is just the change in expected maximum utility converted into a money metric by multiplying by the inverse of the constant marginal utility of money, $(1/\mu)$.[41]

In addition, if there are zero income effects, $CV^R = E(CV) = CV^{med} = E(EV) = EV^R = EV^{med}$. For details see Hanemann (1985) or McFadden(1996).[42] When there are zero income effects, the easiest way to calculate E(CV) is to calculate CV^R using Equation (54).

When there are *step income effects*, and the policies that are under consideration have welfare effects small enough so that the marginal utility of money at $(I - p_{mj} - c)$ is the same as at $(I - p_{mj})$, Equation (54) still applies. With broad steps in the marginal utility of money function, the researcher can estimate compensating variations that vary by income category without subjecting herself to the complication described next.[43]

---

[41]Note that when there are *zero income effects*, if one incorrectly multiples the prices of all the alternatives by $\lambda > 0$ and re-estimates the model, the value of the likelihood function with not change and expected maximum utility will not change, but the estimated marginal utility of money will be wrong, and using Equation (54), one will incorrectly estimate λCV^R rather than CV^R. That is, making alternative prices twice (half) as large as they should be will incorrectly double (halve) the CV estimates

[42]It follows that when there are zero income effect, CV^{med} cannot be the largest or smallest CV, unless all CV are equal.

[43]For example, Morey *et al.* (1997b) found a *step income effects* model more appropriate than either a *zero* or *continuous income effects* model, and from it derive estimates of willingness-to-pay for monument preservation that are significantly lower for the poor. Work by Sharma *et al.* (1997) on the

(continued...)

When there are *continuous income effects*, $CV^R \neq EV^R$, Equation (45) cannot be solved to obtain a closed form solution for the CV^R and Equation (46) cannot be solved to obtain a closed form solution for the EV^R However, it is still easy to numerically calculate an individual's CV^R, or EV^R, for any proposed change. For example, given the estimated parameter values, an individual's CV^R for any proposed change can be calculated by using any numerical minimization algorithm to find the CV^R that minimizes

$$[E(U(y^0,\ P^0,\ \beta^0,\ G,\ <\epsilon_{mj}>))\ -\ E(U(y^1\ -\ CV^R,\ P^1,\ \beta^1,\ G.\ <\epsilon_{mj}>))]^2$$

where E(U) is defined by Equation (41).[44] The CV^R that minimizes this expression will be the individual's CV^R for the change.

Hanemann (1985) and McFadden (1996) demonstrate that when there are continuous income effects $CV^R \neq E(CV)$, and Hanemann demonstrates that CV^R can be greater than or less than E(CV).

If one wants to approximate E(CV) and there are continuous income effects, there is the issue of how best to approximate it. One approximation is CV^R.

With effort, one can more accurately approximate E(CV). It is possible to randomly draw an $<\epsilon_{mj}>$ vector from the GEV distribution (Equation (8)) and then calculate the individual's CV conditional on the drawn $<\epsilon_{mj}>$ being the individual's true epsilons. That is, for the drawn $<\epsilon_{mj}>$, find the *c* that solves Equation (48).[45] An approximation to the E(CV) can be obtained by calculating CV for each of T random draws from the GEV distribution and taking the average. Denote this approximation, $\hat{E}(CV)$. As $T \rightarrow \infty$, $\hat{E}(CV) \rightarrow E(CV)$. An approximation to CV^{med} $(\hat{CV}^{med})$ is the median of these T CV estimates.

---

[43](...continued)
choice of health care providers in Nepal indicates that the choice of provider depends on whether one is poor, but does not vary continuously with income.

[44]For example one can use the *optimization* routine in the statistical package *Gauss* or the *FindMinimum* command in the mathematical software *Mathematica*.

[45]In general, there will not be a closed-form solution for c. However, given $<\epsilon_{mj}>$, *c* can be estimated by using the optimization procedure in *Gauss* (*Optmum*) to search for the *c* that minimizes $M = [U^1(c) - U^0]^2$.

The steps in of calculating $\hat{E}(CV)$ are the same for both the NL model and its special case the ML model, but are much easier to implement for the ML model. This is because the ML model requires only that one make independent draws from the univariate extreme value distribution, but the NL model requires that one draw *C* dimensional vectors from a GEV distribution. The latter is much more difficult.

Consider first a *C* alternative ML model with continuous income effects. Let $<\epsilon_{mj}^{\tau}> \equiv (\epsilon_{11}^{\tau},....\epsilon_{C}^{\tau})$, τ = 1,2,...T, where each ϵ_{mj}^{τ} is a separate draw from the univariate extreme value distribution [Equation (25)]. The vector $<\epsilon_{mj}^{\tau}>$ is constructed by taking *C* such draws. Given $<\epsilon_{mj}^{\tau}>$, calculate CV^{τ} for draw τ. CV^{τ} is the individuals exact compensating variation if $<\epsilon_{mj}^{\tau}>$ is the epsilon vector that the individual actually experiences during that choice occasion. Calculate CV^{τ} for each of the $<\epsilon_{mj}^{\tau}>$, and approximate E(CV) with $\hat{E}(CV) = (1/\mathrm{T})\sum_{\tau=1}^{\mathrm{T}} CV^{\tau}$. Practically speaking, the number of draws, *T*, should be sufficiently large so that $(1/\mathrm{T})\sum_{\tau=1}^{\mathrm{T}} CV^{\tau}$ does not significantly change if *T* is increased. This can be a large number.

For the ML model with continuous income effects, implementing this procedure requires a method for taking random draws from the univariate extreme value distribution. It is well known that $y = F(x) \sim u(0,1)$, where F(x) is the cumulative density function for any scalar random variable x, and *u(0,1)* is the uniform distribution on the unit interval. That is, the distribution of any univariate CDF is uniform on the 0-1 interval. Consider the inverse function, $x = F^{-1}(y)$. If ξ is a random draw from *u(0,1)*, $\epsilon = F^{-1}(\xi)$ is a random draw from $F(\epsilon)$. Since *Gauss*, *Mathematica*, and other math/stat packages have built-in commands for taking random draws from the univariate distribution, one can use $\epsilon = F^{-1}(\xi)$ to generate a random draw from any univariate CDF. Given the extreme value CDF [Equation (25)], $\epsilon = F^{-1}(\xi) = -\ln(-\ln(\xi))$, and $-\ln(-\ln(\xi))$ is a random draw from the extreme value distribution if ξ is a random draw from the unit interval uniform distribution.

In contrast, for a NL model, to calculate $\hat{E}(CV)$ one must draw *C* dimensional epsilon vectors from the GEV distribution [Equation (8)]. McFadden (1996) has developed a technique to do this, but implementation is not for the faint of heart. The technique is complicated and computer intensive because it must account for the correlations across the alternatives implied by the nesting structure.

Herriges and Kling (1997) consider whether calculating $\hat{E}(CV)$, rather than CV^R, is worth the extra effort when one's intent is to estimate E(CV). To investigate this issue, they calculate $\hat{E}(CV)$ for three different policy scenarios for nine different estimated models. For each model, D = 1000. The models vary in terms of the nesting structure and in the form income enters the $<V_{mj}>$. The application is sport fishing and seems to be the first calculation of $\hat{E}(CV)$ for an estimated NL model with continuous income effects. They compare $\hat{E}(CV)$ and CV^R for an increase in catch rates for two different models; both models have continuous income effects; one is ML and the other NL. These comparisons are made at the averages over the sample and by income quartiles. For the NL model, the average of the sample CV^R is \$16.15 and the $\hat{E}(CV)$ is \$16.95, whereas for the ML model they are both \$17.41. The estimated CV^R and $\hat{E}(CV)$, which vary by income, also match closely by income quartiles. None of the differences are statistically significant. Although one should be hesitant to draw general conclusions from specific examples, the results from Herriges and Kling demonstrate that CV^R can closely approximate $\hat{E}(CV)$. Hanemann (1985) also has a numerical example where the CV^R closely approximates E(CV).

In contrast, McFadden (1996) has demonstrated that the bias in CV^R as an estimate of E(CV) can be larger than in the above examples if the policy causes very large changes in utility. He constructs a simple two-alternative example where it is possible to calculate, for a simple change, both the E(CV) and the CV^R. The example has no prices and one site characteristic, where the level of the characteristic at the second site is held at zero. The model is ML and the budget, I, enters the two conditional indirect utility functions as $\sqrt{I}$, where I = 1. For an increase in the site characteristic at site one from zero to .1, the bias in CV^R is 2.5%, for an increase from zero to .4, the bias in CV^R is 10.4%, and for an increase from zero to 1.0, the bias in CV^R is 29.9%.

McFadden (1996) also calculates $\hat{E}(CV)$ for three, three-alternative NL models, all with zero income effects. He chose models with zero income effects so E(CV) could be calculated exactly and compared to the $\hat{E}(CV)$. The three models differ in terms of the assumed value for s. For s = 1 (a ML model), 755 draws were needed to reduce the error in $\hat{E}(CV)$ to less than 5%; for s = 2 , 3206 draws were required; and for s = 10, 18,820 draws were required. These results indicate that fewer draws will be needed to accurately approximate E(CV) with $\hat{E}(CV)$ when the model with income effects is ML rather than NL.

Given the difficulties associated with calculating $\hat{E}(CV)$, McFadden (1996) has developed bounds on E(CV) that can be estimated without taking draws from a GEV distributions. For details see, McFadden (1996), Herriges and Kling (1997) and the comments above on calculating the range of the CV.

In summary, CV^R, $E(CV)$, and CV^{med} are measures of central tendency for $f(CV)$; $E(CV)$ is the mean of the distribution and CV^{med} is the median. Remember that generally there will be a different *f(CV)* for each type of individual. The three measures of central tendency are equal when there are zero income effects and often equal when there are step income effects. When there are continuous income effects, CV^R and $\hat{E}(CV)$ both approximate E(CV), and $\hat{E}(CV)$ can be made as accurate as needed by basing it on a large enough number of draws. $C\hat{V}^{med}$ can be calculated when calculating $\hat{E}(CV)$.

Since these three measures of central tendency are functions of the estimated parameters values and the parameter estimates are random variables with a variance-covariance matrix, these three measures of central tendency are also random variables from the researcher’s perspective, so it is often important to estimate, for each type of individual, a confidence interval on the desired measure. Details on how to do this are provided in *Diversion 6*.

Recall that CV is not the per year (or per season) compensating variation, but rather the compensating variation per choice occasion. If the year is divided into a fixed number of choice occasions (e.g. weeks), the choice set includes all possible alternatives, $<V^0_{mj}>$ does not vary across choice occasions, and $<V^1_{mj}>$ do not vary across choice occasions, the compensating variation for the year (season) is easily obtained by multiplying the per-period CV by the number of periods in the year (season).[46] Factors that might cause

Diversion 6: Estimating the confidence interval on CV^R (or $E(\hat{CV})$ or CV^{med}). One, for example, typically estimates CV^R at the estimated parameter values (the mean values of the estimated variance-covariance matrix). A 95% confidence interval for CV^R is obtained by taking a large number of random draws from the estimated variance-covariance matrix, calculating CV^R for each of these draws to form a distribution of the CV^R estimates, and then forming the 95% confidence interval by deleting the top and bottom 2.5% of the distribution. Note that even if the number of draws is very large, the mean of the CV^R estimates will not, in general, equal the CV^R calculated at the estimated parameters. This is due to the nonlinearity of the CV function - typically they are very similar. .

One randomly draws a parameter vector as follows: Let $\hat{b}$ be the *p x 1* estimated parameter vector, where *cov* is its *p x p* estimated variance-covariance matrix. A randomly drawn parameter vector is $\hat{b}_r = \hat{b} + srcov^{/}rndn(p,1)$, where $rndn(p,1)$ is a *p x 1* random draw from a standard normal and *srcov* is the Cholesky decomposition of *cov*. The number of draws should be sufficiently large so that increaseing the number of draws does not change the confidence interval.

---

[46]Note that alternatives can be grouped. For example, the complete set of alternatives facing the individual could be lumped in two categories, fishing trips and nonfishing trips, where nonfishing trips include staying at home or going bowling. For example, Morey *et al.* (1993) have a 50 choice occasion, nine alternative model where eight of the alternatives are salmon fishing sites and the ninth alternative is
(continued...)

variation in $<V_{mj}>$ across choice occasions include weather, temperature, and catch rates. If the $<V_{mj}^0>$ and $<V_{mj}^1>$ vary across choice occasions, one has to estimate CV for each choice occasion, and then sum across choice occasions.

Alternatively, if the choice set is restricted (does not include all alternatives), things are more complicated and one cannot get the compensating variation for the year by simply multiplying the compensating variation per choice occasion by the number of choice occasions in the year. Models of site choice only have restricted choice sets. The problem is that when the choice set is restricted, the number of choice occasions in the year becomes a function of the attributes of the alternatives in the restricted choice set. Multiplying the CV by the number of choice occasions in the initial state, $\{I^o, P^o, \beta^o\}$, provides only a lower bound on the compensating variation for the year. Multiplying the CV by the number of choice occasions in the proposed state, $\{I^1, P^1, \beta^1\}$, provides only an upper bound on the compensating variation for the year. Neither is necessarily a close approximation. Details are provided in Morey (1994). See also Herriges *et al.* (1997).

---

[46](...continued)
not taking a salmon fishing trip. Morey *et al.* (1995 and 1997) have a 60 choice occasion model where in each period the angler chooses among 32 alternatives: 26 river segments, five regions, and not river fishing in Montana. There is a common presumption that the more aggregated the treatment of other alternatives, the larger, in absolute value, will be the estimated CV for a change in the disaggregated group of interest (see, for example, Montgomery and Needelman,(1997). There is no theoretical foundation for this conjecture; aggregating alternatives is different from excluding alternatives from the choice set. Plantinga *et al.* (1997) demonstrate the conjecture to be false with some empirical counterexamples.

For example: In the **model of participation and site choice**, the year is divided into a finite number of periods and each period includes all alternatives, so CV is the per-period compensating variation and the compensating variation for the year is easily obtained by multiplying the CV by the number of periods.

In the **model of site choice with saltwater sites, lakes, and rivers**, choice occasions are trips and on each trip the individual fishes (staying home is not an option). In this case, CV is the compensating variation per trip. Define $trips^0$ as the predicted number of trips in the initial state and $trips^1$ as the predicted number of trips in the proposed state. CV multiplied by $trips^0$ is a lower bound on the yearly CV. CV multiplied by $trips^1$ is an upper bound.

VII. Expanding The Nest

The NL model can be expanded into as many levels as one desires. For a three-level nest, Equation (1) expands to

(1-3level) $$U_{lmj} = V_{lmj} + \epsilon_{lmj} \qquad \forall\ (lmj) \in C,$$

where the l dimension of the choice set is characterized in term of L distinction types. As before, the m dimension has M distinct types and the j dimension has J distinct types. To generate a simple three-level NL model where $s_m = s\ \forall m$ and $a_m = 1\ \forall m$, assume[47]

(8-3level-1) $$F(<\epsilon_{lmj}>) = \exp\{-\sum_{l=1}^{L}[\sum_{m=1}^{M_l}[\sum_{j=1}^{J_{lm}} e^{-s\epsilon_{lmj}}]^{(t/s)}]^{1/t}\}.$$

If $t = s$, this collapses to a two-level nest. The derivation of the *Prob(gni)* and expected maximum utility follow the same logic as in the two-level case; the notation is just messier. If one bashes through the derivations, one obtains

[47] A necessary and sufficient condition for this function to be globally well-behaved is $s \geq t \geq 1$.

(16-3level-1) $$Prob(gni) = \frac{e^{sV_{gni}} \; [\sum_{m=1}^{M_g} [\sum_{j=1}^{J_{gm}} e^{sV_{gmj}}]^{(t/s)}]^{(1/t)-1} \; [\sum_{j=1}^{J_{gn}} e^{sV_{gnj}}]^{((t/s)-1)}}{\sum_{l=1}^{L} [\sum_{m=1}^{M_l} [\sum_{j=1}^{J_{lm}} e^{sV_{lmj}}]^{(t/s)}]^{1/t}}$$

and per-choice occasion expected maximum utility is

(41) $U = \ln D + .57721....$,

where

(35-3level-1) $$D = \sum_{l=1}^{L} [\sum_{m=1}^{M} [\sum_{j=1}^{J_m} e^{sV_{lmj}}]^{(t/s)}]^{1/t} .$$

One example of a three-level nest: A **model of participation and site choice for Atlantic Salmon fishing** [Morey *et al.* (1993] with C = 9 alternatives (staying at home and visiting one of eight salmon rivers). Assume L has two elements: 1= going salmon fishing and 0 = not going salmon fishing. M_1 has two element: 1= fishing a river in Maine and 2 = fishing a river in Canada. J_1 includes five rivers in Maine and J_2 includes three rivers in Canada. A three-level nest seems appropriate because one would expect the random components in the conditional indirect utility functions for the Maine rivers to be more correlated with each other than they are with the random components in the conditional indirect utility functions for the Canadian rivers. One would also expect the random components in the conditional indirect utility functions for the rivers to be more correlated with each other than they are with the random component in the conditional indirect utility functions for nonparticipation.

Equation (8-3level) takes the form

$F(<\epsilon_{lmj}>) = \exp\{-\exp(-\epsilon_0) - [\,[\exp(-s\epsilon_{11})+...+\exp(-s\epsilon_{15})]^{t/s} + [\exp(-s\epsilon_{21})+...$
$\exp(-s\epsilon_{23})]^{t/s}\,]^{1/t}\,\}$

and the probability of choosing the first river in Canada is

$Prob(121) = \exp(sV_{121})\,[\,[\exp(sV_{111})+...+\exp(sV_{115})]^{t/s}$
$+ [\exp(sV_{121})+...+\exp(sV_{123})]^{t/s}\,]^{1/t-1}$ multiplied by $[\exp(sV_{121})+...+\exp(sV_{123})]^{t/s-1}$

and divided by

$\{\exp(V_0) + [\,[\exp(sV_{111})+...+\exp(sV_{115})]^{t/s} + [\exp(sV_{121})+...+\exp(sV_{123})]^{t/s}\,]^{1/t}$

```
@ nstl-3lv.cmd   June 30, 92  @

@this is the gauss program that was used to estimate the three level NL model of Atlantic Salmon
Fishing that was described in the previous box.  For more details see Morey et al. (1993) @

library maxlik;  @ These three commands load the maximum likelihood module in Gauss @
#include maxlik.ext;
maxset;

dataset = "mst4lgt";   @ the data is stored in the Gauss data set "mst4lgt"  @
output file = nstl-3lv.out on;

proc li(b,x); @ "li" is the procedure that generates the ln of the lik function.  "li" is called below by the
        maximum liklihood module, "maxlik" @

local ppy, evp, evd, evm, evk, evs, evns, evnb, evq, evn, vp, vm, vd, vk, vs, vns, vnb, vq, vn,
        inclusm, inclusc, inclusp, inclus, linclus, lsump, lmsum, lcsum, x;

ppy = x[.,4]/50; @ x[.,4] is income @

@ the following are the exp of the conditional indirects with the
conditional indirects for the site alternatives multiplied by s;i.e b[11].  Note that the conditional indirect
for nonparticipation, n, is not multiplied by s because it cancels out. @

@ In the following b[4] is a constant term, b[11]=s, b[12] = t, x[.,1] = years salmon fishing. x[.,2] = 1 if
member of a fishing club and zero otherwise, x[.,3] = individual's age, x[.,5] = number of periods
individual did not salmon fish,  x[.,6]- x[.,13] are the expected catch rates at the eight sites, x[.,14] -
x[.,21]are trip costs for the eight sites, and x[.,22]- x[.,29] are the number of trips each individual took to
each of the eight sites, @

 evp = exp(b[11]*(b[1]*(ppy-x[.,14])+b[2]*x[.,6]+b[3]*x[.,6]^.5
    +b[10]*((1728.4720+ppy-x[.,14])^.5)));
evm = exp(b[11]*(b[1]*(ppy-x[.,15])+b[2]*x[.,7]+b[3]*x[.,7]^.5
    +b[10]*((1728.4720+ppy-x[.,15])^.5)));
                     evd = exp(b[11]*(b[1]*(ppy-x[.,16])+b[2]*x[.,8]+b[3]*x[.,8]^.5
                         +b[10]*((1728.4720+ppy-x[.,16])^.5)));
                     evk = exp(b[11]*(b[1]*(ppy-x[.,17])+b[2]*x[.,9]+b[3]*x[.,9]^.5
                         +b[10]*((1728.4720+ppy-x[.,17])^.5)));
                     evs = exp(b[11]*(b[1]*(ppy-x[.,18])+b[2]*x[.,10]+b[3]*x[.,10]^.5
                         +b[10]*((1728.4720+ppy-x[.,18])^.5)));


                     @the program is continued in the next box @
```

```
@continuation of previous box @

evns= exp(b[11]*(b[1]*(ppy-x[.,19])+b[2]*x[.,11]+b[3]*x[.,11]^.5
     +b[10]*((1728.4720+ppy-x[.,19])^.5)));
evnb= exp(b[11]*(b[1]*(ppy-x[.,20])+b[2]*x[.,12]+b[3]*x[.,12]^.5
     +b[10]*((1728.4720+ppy-x[.,20])^.5)));
evq = exp(b[11]*(b[1]*(ppy-x[.,21])+b[2]*x[.,13]+b[3]*x[.,13]^.5
     +b[10]*((1728.4720+ppy-x[.,21])^.5)));

@the next line is the exp of the condit indirect for nonpartic @

evn = exp(b[1]*ppy+b[4]+b[5]*x[.,1]+b[6]*x[.,2]+b[7]*x[.,3]
     +b[8]*x[.,1]^.5+b[9]*x[.,3]^.5+b[10]*((1728.4720+ppy)^.5));

vp=ln(evp); vm=ln(evm); vd=ln(evd); vk=ln(evk); vs=ln(evs);vns=ln(evns); vnb=ln(evnb); vq=ln(evq);
vn=ln(evn);

@ Note b[12] is t @

inclusm = (evp+evm+evd+evk+evs)^(b[12]/b[11]);
inclusc = (evns+evnb+evq)^(b[12]/b[11]);
inclusp = (inclusm + inclusc)^(1/b[12]);

@ inclus is the denomin in all the prob @

inclus  = evn + inclusp;
linclus=ln(inclus);
lsump = ((1/b[12])-1)*ln(inclusm+inclusc);
lmsum = ((b[12]/b[11])-1) * ln(evp+evm+evd+evk+evs);
lcsum = ((b[12]/b[11])-1) * ln(evns+evnb+evq);

@ the next command calculates the contribution to the log of the lik function for each indiv in the
sample @

retp(x[.,22].*(vp + lsump + lmsum - linclus)
     + x[.,23].*(vm + lsump + lmsum - linclus)
+ x[.,24].*(vd + lsump + lmsum - linclus)
     + x[.,25].*(vk + lsump + lmsum - linclus)
     + x[.,26].*(vs + lsump + lmsum - linclus)
     + x[.,27].*(vns+ lsump + lcsum - linclus)
     + x[.,28].*(vnb+ lsump + lcsum - linclus)
     + x[.,29].*(vq + lsump + lcsum - linclus)
     + x[.,5].*(vn - linclus));
endp;

@ continued in the next box @
```

```
@continuation from previous box @

@ the following are the converged values from the 6/30/92 run - the AJAE estimates @

startv = {.002190, -1.729160, 5.912200, 8.850202, .095329, -.805250,.170146, -1.227537, -1.856089,
1.088725, 1.307125, .611724};

_title = "NSTL-3LEV INCOME EFFECTS ";
_mlmiter = 2000;
_mlgtol  = .0001;
{bbb,f0,g,h,retcode}=maxprt(maxlik("mst4lgt",0,&li,startv));

bmm94 = bbb;     save bmm94;    output off;
```

In conclusion, consider a more general three-level NL model, where[48]

(8-3level-2) $$F(<\epsilon_{lmj}>) = \exp\{-\sum_{l=1}^{L} x_l \, [\sum_{m=1}^{M} a_m^{t_l} \, [\sum_{j=1}^{J_m} e^{-s_m \epsilon_{lmj}}]^{(t_l/s_m)}]^{1/t_l}\}$$

$$= \exp\{-\sum_{l=1}^{L} \, [\sum_{m=1}^{M} \, [\sum_{j=1}^{J_m} e^{-s_m(\epsilon_{lmj} - \alpha_m - \chi_l)}]^{(t_l/s_m)}]^{1/t_l}\}$$

where $a_m = e^{\alpha_m}$, and $x_l = e^{\chi_l}$. In which case,

(16-3level-2) $$Prob(gni) = \frac{e^{s_n V_{gni}} x_g a_n^{t_g} \, [\sum_{m=1}^{M_g} a_m^{t_g} [\sum_{j=1}^{J_{gm}} e^{s_m V_{gmj}}]^{(t_g/s_m)}]^{(1/t_g)-1} \; [\sum_{j=1}^{J_{gn}} e^{s_n V_{gnj}}]^{((t_g/s_n)-1)}}{\sum_{l=1}^{L} x_l [\sum_{m=1}^{M_l} a_m^{t_l} [\sum_{j=1}^{J_{lm}} e^{s_m V_{lmj}}]^{(t_l/s_m)}]^{1/t_l}}$$

[48]Given $a_m > 0 \;\forall m,\; x_l > 0 \;\forall l$, the necessary and sufficient conditions for this density function to be globally well behaved are $\{s_m \geq t_l \geq 1 \;\; \forall m \in M_l\},\; l=1,2,...L.$

$$= \frac{e^{s_n(\chi_g+\alpha_n+V_{gni})} \; [\sum_{m=1}^{M_g} [\sum_{j=1}^{J_{gm}} e^{s_m(\chi_g+\alpha_m+V_{gmj})}]^{(t_g/s_m)}]^{(1/t_g)-1} \; [\sum_{j=1}^{J_{gn}} e^{s_n(\chi_g+\alpha_n+V_{gnj})}]^{((t_g/s_n)-1)}}{\sum_{l=1}^{L} [\sum_{m=1}^{M_l} [\sum_{j=1}^{J_{lm}} e^{s_m(\chi_l+\alpha_m+V_{lmj}}]^{(t_l/s_m)}]^{1/t_l}} .$$

The $<\chi_l>$ add a group specific constant term to each of the V_{lmj} for each of the L groups, and the $<\alpha_m>$ add a group specific constant term to each of the V_{lmj} for each of the M groups,

(35-3level-2) $$D = \sum_{l=1}^{L} x_l \; [\sum_{m=1}^{M_l} a_m^{t_l} \; [\sum_{j=1}^{J_{lm}} e^{s_m V_{lmj}}]^{(t_l/s_m)}]^{1/t_l} \; .$$

References

W.J. Adamowicz, J. Swait, P. Boxall, J. Louviere and M. Williams, "Perceptions verus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation," paper presented at the 1996 AERE workshop, Combining Stated and Revealed Preference Data to Estimate the Demand for and Benefits from Environmental Amenities, Boulder CO (1996).

W.J. Adamowicz, J Louviere and M. Williams, "Combining Stated and Revealed Preference Method for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, **26**, 271-292 (1994).

T. Amemiya, "On a Two-Step Estimation of a Multivariate Logit Model," *Journal of Econometrics*, **8**, 12-21 (1978).

S.P Anderson, A De Palma and J Thisse, "The CES is a Discrete Choice Model," *Economic Letters*, **23**, 139-140 (1987).

________ "Discrete Choice Theory of Product Differentiation," MIT Press, Cambridge Massachusetts (1992).

N.E. Bockstael, W.M. Hanemann and C.L. Kling, "Modeling Recreational Demand in a Multiple Site Framework," *Water Resources Research*, **23**, 951-60 (1987).

N.E. Bockstael, W.M. Hanemann and I.E. Strand, "Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models," Report prepared for the Environmental Protection Agency under Cooperative Agreement CR-811043-01-0, Washington D.C. (1986).

N.E. Bockstael, K.E. McConnell and I.E. Strand, "A Random Utility Model of Sport Fishing: Some Preliminary Results for Florida," *Marine Resource Economics*, **6**, 245-260 (1989).

___________ "Recreation, in "Measuring the Demand for Environmental Quality," (John B. Braden and Charles Kolstad, eds.), New York, North Holland (1991).

D. Brownstone and K.A. Small, "Efficient Estimation of Nested Logit Models," *Journal of Business and Economic Statistics*, **7**, 67-74 (1989).

A. Börsch-Supan, "On the Compatibility of Nested Logit Models with Utility Maximization," *Journal of Econometrics*, **43**, 373-388 (1990).

T. Buchanan, E.R. Morey and D.M. Waldman, "♬Happy Trails to You: The Impact of Trail Characteristics and Access Fees on a Mountain Biker's Trail Selection and Consumer's Surplus," discussion paper, Department of Economics, University of Colorado-Boulder, (1997).

T.A. Cameron, Qualitative Choice Modeling of Energy Conservation Decisions, unpublished Ph.D. Dissertation, Princeton University, Department of Economics (1982).

___________, "A Nested Logit Model of Energy Conservation Activity by Owners of Existing Single Family Dwellings," *The Review of Economics and Statistics*, **55**, 205-211 (1985).

R. Carson, W.M. Hanemann and T Wegge, "Southcentral Alaska Sport Fishing Study," Report prepared by Jones and Stokes Associates for Alaska Department of Fish and Game, Anchorage, AK 99502 (1987)

P. Caulkins, R. Bishop and N. Bouwes, "The Travel Cost Model for Lake Recreation: A Comparison of Two Methods for Incorporating Site Quality and Substitution Effects," *American Journal of Agricultural Economics*, **68**, 291-297 (1986).

H.Z. Chen, "Angler's Perceptions of Environmental Quality and Benefit Estimation in Multinomial Probit Model: A Simulation Approach, paper presented at the 1996 AERE workshop, Combining Stated and Revealed Preference Data to Estimate the Demand for and Benefits from Environmental Amenities, Boulder CO (1996).

H.Z. Chen, F. Lupi and J. Hoehn, "An Empirical Assessment of Multinomial Probit and Logit Models of Recreation Demand," Forthcoming as Chapter 8 in "Valuing the Environment Using Recreation Demand Models," (C.L. Kling and J.A. Herriges, Eds.), Edward Elgar Publishing Ltd. (1997).

M. Creel and J. Loomis, "Recreation Value of Water to Wetlands in the San Joaquin Valley: Linked Multinomial Logit and Count Data Trip Frequency Models," *Water Resources Research*, **28**, 2597-2606 (1992).

D. Feenburg and E. Mills, "Measuring the Benefits of Water Pollution Abatement," Academic Press, New York (1980).

A.B. Hauber and G.R. Parsons, "Nesting Structure Choice in a Random Utility Model of Recreation Demand," discussion paper, Department of Economics, University of Delaware, (October 1996).

J.A. Hausman, G.K. Leonard and D McFadden, "A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage," *Journal of Public Economics*, **56**, 1-30 (1995).

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

_________, "Welfare Analysis with Discrete Choice Models," working paper, Department of Resource and Agricultural Economics, University of California, Berkeley, (March 1985).

D.A. Hensher, "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model," *The Review of Economics and Statistics*, **56**, 657-667 (1986).

J.A. Herriges and C.L. Kling, "Nonlinear Income Effects in Random Utility Models,"discussion paper, Department of Economics, Iowa State University, (June 4, 1997).

_________, "Testing the Consistency of Nested Logit Models with Utility Maximization," *Economic Letters*, **50**, 33-39 (1996).

J.A. Herriges, C.L Kling and D.J. Phaneuf, "Corner Solution Models of Recreation Demand: A Comparison of Competing Frameworks," Forthcoming as Chapter 9 in Valuing the Environment Using Recreation Demand Models," (C.L. Kling and J.A. Herriges, Eds.) Edward Elgar Publsihing Ltd. (1997).

J.P, Hoehn, T. Tomasi, F. Lupi and H.Z. Chen, "An Economic Model for Valuing Recreational Angling Resources in Michigan," Report submitted to the Environmental Response Division, Michigan Department of Environmental Quality, (December 1996).

N.L Johnson, S. Kotz and N Balakrishnan, "Continuous Univariate Distributions - Second Edition," Vols. 1 and 2," Wiley, New York (1994).

C.A. Jones and Y.D. Sung, "Models for Censored Duration Data, Competing Risks, and Time-Varying Covariates: with an Application to Recreational Trip Demand," discussion paper, NOAA, US Department of Commerce (1992).

C.L. Kling and J.A. Herriges, "An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization," *American Journal of Agricultural Economics*, **77**, 875-884 (1995)

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

D.F. Layton, "Rank Ordered, Random Coefficients Multinomial Probit Models for Stated Preference Surveys: An Investigation of Preference Heterogeneity, IIA and Contingent Ranking Data," paper presented at the 1996 AERE workshop, Combining Stated and Revealed Preference Data to Estimate the Demand for and Benefits from Environmental Amenities, Boulder CO (1996).

C.F. Manski and S.R. Lerman, "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, **45**, 1977-1988 (1977.

D. McFadden, Modeling the Choice of Residential Location, Chapter 3 in "Spatial Interaction Theory and Planning Models" (A. Karlqvist, L, Lundqvist, F. Snickars and J Weibull, Eds.) North-Holland, Amsterdam (1978).

________, Econometric Models of Probabilistic Choice, Chapter 5 in "Structural Analysis of Discrete Data with Econometric Applications" (C. Manski and D. McFadden eds.), MIT Press, Cambridge (1981).

________, "A Method of Simulated Moments For Estimation: Discrete Response Models Without Numerical Integration," *Econometrica*, **57**, 995-1026 (1989).

________, "Computing Willingness-to-Pay in Random Utility Models," discussion paper, Department of Economics, University of California, Berkeley, (1996).

J.W. Milon, "A Nested Demand Shares Model of Artificial Marine Habitat Choice by Sport Anglers," *Marine Resources Research*, **5**, 191-213 (1988).

M. Montgomery and M. Needelman,"The Welfare Effects of Toxic Contamination in Freshwater Fish," Land Economics, 73, 211-223 (1997).

E.R. Morey, "The Demand for Site-Specific Recreational Activities: A Characteristics Approach," *Journal of Environmental Economics and Management*, 8, 245-271 (1981).

________, "What is Consumer's Surplus Per Day of Use, When is it a Constant Independent of the Number of Days of Use, and What Does It Tell Us About Consumer's Surplus?" *Journal of Environmental Economics and Management* **26**, 257-270 (1994).

E.R. Morey, W.S. Breffle, R.D. Rowe and D.M. Waldman, "Revised Report and Rebuttal: Assessment of Damages to Anglers and Other Recreators from Injuries to the Upper Clark Fork River Basin," Report prepared by Hagler Bailly Consulting for the State of Montana, Natural Resource Damage Litigation Program, (October 18, 1995).

________, "Estimating the Damages for Injuries to Trout in Montana's Clark Fork River: Participation, Site Choice, and Endogenous Expected Catch Rates, discussion paper, Department of Economics, University of Colorado, Boulder (1997a).

E.R. Morey and K. Rossmann, "Estimating Individual-Specific Scale Parameters in Logit Models of Hypothetical Choice," ongoing research, Department of Economics, University of Colorado, 1997.

E.R. Morey, K. Rossmann, L Chestnut, and S Ragland, "Valuing Acid Deposition Injuries to Cultural Resources," A report prepared for the National Acid Precipitation Assessment Program by the University of Colorado and Hagler Bailly Consulting, (May 14, 1997b).

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

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

A. Pakes and D. Pollard, "Simulation And the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1057 (1989).

G.R. Parsons and M.J. Kealy, "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation," *Land Economics*, **68**, 93-106 (1992).

________, "A Demand Theory for Number of Trips in a Random Utility Model of Recreation," *Journal of Environmental Economics and Management*, **29**, 357-367 (1995).

G.R. Parsons and M.S. Needleman, "Site Aggregation in a Random Utility Model of Recreation," *Land Economics*, **68**, 418-433 (1992).

A.J. Plantinga, G.R. Parsons and K.J. Boyle, "Focused Choice Sets and Targeted Aggregation in Random Utility Model," discussion paper, Department of Economics, University of Maine (1997).

V. Sharma, E.R. Morey and A. Mills, "Estimating Compensating Variations of Malaria Patients for Health Care Proposals in Nepal," discussion paper, Department of Economics, University of Colorado, Boulder, (1997).

V.K. Smith and R. Von Haefen, "Welfare Measurement and Representative Consumer Theory," discussion paper, Department of Economics, Duke University (1997).

J. Swait and W. Adamowicz, "The Effect of Choice Environment and Task Demand on Consumer Behavior: Discriminating Between Contribution and Confusion, " paper presented at the 1996 Canadian Resource and Environmental Economics Study Group, Montreal (1996).

F. Verboven, "The Nested Logit Model and Representative Consumer Theory," *Economic Letters*, **50**, 57-63 (1996).