CHAPTER 5: Flexible Model Structures for Discrete Choice Analysis

Chandra R. Bhat *
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu
University of Texas at Austin

Naveen Eluru
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: NaveenEluru@mail.utexas.edu

and

Rachel B. Copperman
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: RCopperman@mail.utexas.edu

* Corresponding author

ABSTRACT

Econometric discrete choice analysis is an essential component of studying individual choice behavior. In this chapter, we provide an overview of the motivation for, and structure of, advanced discrete choice models derived from random-utility maximization.

1 INTRODUCTION

Econometric discrete choice analysis is an essential component of studying individual choice behavior and is used in many diverse fields to model consumer demand for commodities and services. Typical examples of the use of econometric discrete choice analysis include studying labor force participation, residential location, and house tenure status (owning versus renting) in the economic, geography, and regional science fields, respectively; choice of travel mode, destination and car ownership level in the travel demand field; purchase incidence and brand choice in the marketing field; and choice of marital status and number of children in sociology.

In this chapter, we provide an overview of the motivation for, and structure of, advanced discrete choice models derived from random-utility maximization. The discussion is intended to familiarize readers with structural alternatives to the multinomial logit (MNL) and to the models discussed in Chapter 13. Before proceeding to a review of advanced discrete choice models, the assumptions of the MNL formulation are summarized. This is useful since all other random-utility maximizing discrete choice models focus on relaxing one or more of these assumptions.

There are three basic assumptions which underlie the MNL formulation.

The *first assumption* is that the random components of the utilities of the different alternatives are independent and identically distributed (IID) with a type I extreme-value (or Gumbel) distribution. The assumption of *independence* implies that there are no common unobserved factors affecting the utilities of the various alternatives. This assumption is violated, for example, if a decision-maker assigns a higher utility to all transit modes (bus, train, *etc.*) because of the opportunity to socialize or if the decision maker assigns a lower utility to all the transit modes because of the lack of privacy. In such situations, the same underlying unobserved factor (opportunity to socialize or lack of privacy) impacts on the utilities of all transit modes. As indicated in Chapter 13, presence of such common underlying factors across modal utilities has implications for competitive structure. The assumption of *identically distributed* (across alternatives) random utility terms implies that the extent of variation in unobserved factors affecting modal utility is the same across all modes. In general, there is no theoretical reason to believe that this will be the case. For example, if comfort is an unobserved variable the values of which vary considerably for the train mode (based on, say, the degree of crowding on different train routes) but little for the automobile mode, then the random components for the automobile and train modes will have different variances. Unequal error variances have significant implications for competitive structure.

The *second assumption* of the MNL model is that it maintains homogeneity in responsiveness to attributes of alternatives across individuals (*i.e.*, an assumption of response homogeneity). More specifically, the MNL model does not allow sensitivity (or taste) variations to an attribute (*e.g.*, travel cost or travel time in a mode choice model) due to unobserved individual characteristics. However, unobserved individual characteristics can and generally will affect responsiveness. For example, some individuals by their intrinsic nature may be extremely time-conscious while other individuals may be "laid back" and less time-conscious. Ignoring the effect of unobserved individual attributes can lead to biased and inconsistent parameter and choice probability estimates (see Chamberlain, 1980).

The *third assumption* of the MNL model is that the error variance-covariance structure of the alternatives is identical across individuals (*i.e.*, an assumption of error variance-covariance homogeneity). The assumption of identical variance across individuals can be violated if, for example, the transit system offers different levels of comfort (an unobserved variable) on different routes (*i.e.*, some routes may be served by transit vehicles with more comfortable seating and temperature control than others). Then, the transit error variance across individuals along the two routes may differ. The assumption of identical error covariance of alternatives across individuals may not be appropriate if the extent of substitutability among alternatives differs across individuals. To summarize, error variance-covariance homogeneity implies the same competitive structure among alternatives for all individuals, an assumption which is generally difficult to justify.

The three assumptions discussed above together lead to the simple and elegant closed-form mathematical structure of the MNL. However, these assumptions also leave the MNL model saddled with the “independence of irrelevant alternatives” (IIA) property at the individual level [Luce and Suppes (1965); for a detailed discussion of this property see also Ben-Akiva and Lerman (1985)]. Thus, relaxing the three assumptions may be important in many choice contexts.

In this chapter the focus is on three classes of discrete choice models that relax one or more of the assumptions discussed above. The first class of models (labeled as “heteroscedastic models”) is relatively restrictive; they relax the identically distributed (across alternatives) error term assumption, but do not relax the independence assumption (part of the first assumption above) or the assumption of response homogeneity (second assumption above). The second class of models (labeled as “mixed multinomial logit (MMNL) models”) and the third class of models (labeled as “mixed generalized extreme value (MGEV) models”) are very general; models in this class are flexible enough to relax the independence and identically distributed (across alternatives) error structure of the MNL as well as to relax the assumption of response homogeneity. The relaxation of the third assumption implicit in the multinomial logit (and identified on the previous page) is not considered in detail in this chapter, since it can be relaxed within the context of any given discrete choice model by parameterizing appropriate error structure variances and covariances as a function of individual attributes [See Bhat (2007) for a detailed discussion of these procedures.].

The reader will note that the generalized extreme value (GEV) models described in Chapter 13 relax the IID assumption partially by allowing correlation in unobserved components of different alternatives. The advantage of the GEV models is that they maintain closed-form expressions for the choice probabilities. The limitation of these models is that they are consistent with utility maximization only under rather strict (and often empirically violated) restrictions on the dissimilarity and allocation parameters (specifically, the dissimilarity and allocation parameters should be bounded between 0 and 1 for global consistency with utility maximization, and the allocation parameters for any alternative should add to 1). The origin of these restrictions can be traced back to the requirement that the variance of the joint alternatives be identical in the GEV models. Also, GEV models do not relax assumptions related to taste homogeneity in response to an attribute (such as travel time or cost in a mode choice model) due to unobserved decision-maker characteristics, and cannot be applied to panel data with temporal correlation in unobserved factors within the choices of the same decision-making agent. However, GEV models do offer computational tractability, provide a theoretically sound measure for benefit valuation, and can form the basis for formulating mixed models that accommodate random taste variations and temporal correlations in panel data (see Section 4).

The rest of this chapter is structured as follows. The class of heteroscedastic models, mixed multinomial logit models, and mixed generalized extreme value models are discussed in Sections 2, 3, and 4, respectively. Section 5 presents recent advances in the area of simulation techniques to estimate the mixed multinomial and mixed generalized extreme value class of models of Section 3 and 4 (the estimation of the heteroscedastic models in section 2 does not require the use of simulation and is discussed within Section 2). Section 6 concludes the paper with a summary of the growing number of applications that use flexible discrete choice structures.

2 THE HETEROSCEDASTIC CLASS OF MODELS

The concept that heteroscedasticity in alternative error terms (*i.e.*, independent, but not identically distributed error terms) relaxes the IIA assumption has been recognized for quite some time now. Three models have been proposed that allow non-identical random components. The first is the negative exponential model of Daganzo (1979), the second is the oddball alternative model of Recker (1995) and the third is the heteroscedastic extreme-value (HEV) model of Bhat (1995). Of these, Daganzo's model has not seen much application, since it requires that the perceived utility of any alternative not exceed an upper bound (this arises because the negative exponential distribution does not have a full range). Daganzo's model also does not nest the MNL model. Recker (1995) proposed the oddball alternative model which permits the random utility variance of one "oddball" alternative to be larger than the random utility variances of other alternatives. This situation might occur because of attributes which define the utility of the oddball alternative, but are undefined for other alternatives. Recker's model has a closed-form structure for the choice probabilities. However, it is restrictive in requiring that all alternatives except one have identical variance.

Bhat (1995) formulated the HEV model, which assumes that the alternative error terms are distributed with a type I extreme value distribution. The variances of the alternative error terms are allowed to be different across all alternatives (with the normalization that the error terms of one of the alternatives have a scale parameter of one for identification). Consequently, the HEV model can be viewed as a generalization of Recker's oddball alternative model. The HEV model does not have a closed-form solution for the choice probabilities, but involves only a one-dimensional integration regardless of the number of alternatives in the choice set. It also nests the MNL model and is flexible enough to allow differential cross-elasticities among all pairs of alternatives. In the remainder of this discussion of heteroscedastic models, the focus is on the HEV model.

2.1. HEV Model Structure

The random utility of alternative U_i of alternative *i* for an individual in random utility models takes the form (we suppress the index for individuals in the following presentation):

$$U_i = V_i + \varepsilon_i , \qquad (1)$$

where V_i is the systematic component of the utility of alternative *i* (which is a function of observed attributes of alternative *i* and observed characteristics of the individual), and ε_i is the random component of the utility function. Let *C* be the set of alternatives available to the individual. Let the random components in the utilities of the different alternatives have a type I extreme value distribution with a location parameter equal to zero and a scale parameter equal to θ_i for the *i*th alternative. The random components are assumed to be independent, but non-identically distributed. Thus, the probability density function and the cumulative distribution function of the random error term for the *i*th alternative are:

$$f(\varepsilon_i) = \frac{1}{\theta_i} e^{-\varepsilon_i/\theta_i} e^{-e^{-\varepsilon_i/\theta_i}} \text{ and } F_i(z) = \int_{\varepsilon_i=\infty}^{\varepsilon_i=z} f(\varepsilon_i)d\varepsilon_i = e^{-e^{-z/\theta_i}}. \tag{2}$$

The random utility formulation of Equation (1), combined with the assumed probability distribution for the random components in Equation (2) and the assumed independence among the random components of the different alternatives, enables us to develop the probability that an individual will choose alternative *i* from the set *C* of available alternatives

$$\begin{aligned} P_i &= \text{Prob}(U_i > U_j), \quad \text{for all } j \neq i,\, j \in C \\ &= \text{Prob}(\varepsilon_j \leq V_i - V_j + \varepsilon_i), \quad \text{for all } j \neq i,\, j \in C \\ &= \int_{\varepsilon_i=-\infty}^{\varepsilon_i=+\infty} \prod_{j\in C,\, j\neq i} \Lambda\left[\frac{V_i - V_j + \varepsilon_i}{\theta_j}\right] \frac{1}{\theta_i} \lambda\left(\frac{\varepsilon_i}{\theta_i}\right) d\varepsilon_i \end{aligned} \tag{3}$$

where $\lambda(.)$ and $\Lambda(.)$ are the probability density function and cumulative distribution function of the standard type I extreme value distribution, respectively, and are given by (see Johnson and Kotz, 1970)

$$\lambda(t) = e^{-t} e^{-e^{-t}} \text{ and } \Lambda(t) = e^{-e^{-t}}. \tag{4}$$

Substituting $w = \varepsilon_i / \theta_i$ in Equation (3), the probability of choosing alternative *i* can be re-written as follows:

$$P_i = \int_{w=-\infty}^{w=+\infty} \prod_{j\in C,\, j\neq i} \Lambda\left[\frac{V_i - V_j + \theta_i w}{\theta_j}\right] \lambda(w)dw. \tag{5}$$

If the scale parameters of the random components of all alternatives are equal, then the probability expression in Equation (5) collapses to that of the MNL (note that the variance of the random error term ε_i of alternative *i* is equal to $\pi^2\theta_i^2/6$, where θ_i is the scale parameter).

The HEV model discussed above avoids the pitfalls of the IIA property of the MNL model by allowing different scale parameters across alternatives. Intuitively, we can explain this by realizing that the error term represents unobserved characteristics of an alternative; that is, it represents uncertainty associated with the expected utility (or the systematic part of utility) of an alternative. The scale parameter of the error term, therefore, represents the level of uncertainty. It sets the relative weights of the systematic and uncertain components in estimating the choice probability. When the systematic utility of some alternative *l* changes, this affects the systematic utility differential between another alternative *i* and the alternative *l*. However, this change in the systematic utility differential is tempered by the unobserved random component of alternative *i*. The larger the scale parameter (or equivalently, the variance) of the random error component for alternative *i*, the more tempered is the effect of the change in the systematic utility differential (see the numerator of the cumulative distribution function term in Equation 5) and smaller is the elasticity effect on the probability of choosing alternative *i*. In particular, two alternatives will have the same elasticity effect due to a change in the systematic utility of another alternative only if they have the same scale parameter on the random components. This property is a logical and intuitive extension of the case of the MNL, in which all scale parameters are constrained to be equal and, therefore, all cross-elasticities are equal.

Assuming a linear-in-parameters functional form for the systematic component of utility for all alternatives, the relative magnitudes of the cross-elasticities of the choice probabilities of any two alternatives *i* and *j* with respect to a change in the *k*th level of service variable of another alternative *l* (say, x_{kl}) are characterized by the scale parameter of the random components of alternatives *i* and *j*:

$$\begin{aligned} &\eta_{x_{kl}}^{P_i} > \eta_{x_{kl}}^{P_j} && \text{if } \theta_i < \theta_j \\ &\eta_{x_{kl}}^{P_i} = \eta_{x_{kl}}^{P_j} && \text{if } \theta_i = \theta_j \\ &\eta_{x_{kl}}^{P_i} < \eta_{x_{kl}}^{P_j} && \text{if } \theta_i > \theta_j. \end{aligned} \tag{6}$$

2.2. HEV Model Estimation

The HEV model can be estimated using the maximum likelihood technique. Assume a linear-in-parameters specification for the systematic utility of each alternative given by $V_{qi} = \beta' X_{qi}$ for the qth individual and ith alternative (the index for individuals is introduced in the following presentation since the purpose of the estimation is to obtain the model parameters by maximizing the likelihood function over all individuals in the sample). The parameters to be estimated are the parameter vector β and the scale parameters of the random component of each of the alternatives (one of the scale parameters is normalized to one for identifiability). The log likelihood function to be maximized can be written as:

$$L = \sum_{q=1}^{q=Q} \sum_{i \in C_q} y_{qi} \log \left\{ \int_{w=-\infty}^{w=+\infty} \prod_{j \in C_q,\, j \neq i} \Lambda \left[\frac{V_{qi} - V_{gj} + \theta_i w}{\theta_j} \right] \lambda(w) dw \right\}, \tag{7}$$

where C_q is the choice set of alternatives available to the qth individual and y_{qi} is defined as follows:

$$y_{qi} = \begin{cases} 1 \text{ if the } q \text{ th individual chooses alternative } i & (q = 1, 2, \ldots, Q,\ i = 1, 2, \ldots, I) \\ 0 \text{ otherwise}. \end{cases} \tag{8}$$

The log (likelihood) function in Equation (7) has no closed-form expression, but can be estimated in a straightforward manner using Gaussian quadrature. To do so, define a variable. Then, $\lambda(w)dw = -e^{-u}$ and $w = -\ln u$. Also define a function G_{qi} as:

$$G_{qi}(u) = \prod_{j \in C_q,\, j \neq i} \Lambda \left[\frac{V_{qi} - V_{qj} - \theta_i \ln u}{\theta_j} \right] \tag{9}$$

Equation (7) can be written as

$$L = \sum_{q} \sum_{i \in C_q} y_{qi} \, \log \left\{ \int_{u=0}^{u=\infty} G_{qi}(u) \, e^{-u} \, du \right\}. \tag{10}$$

The expression within parenthesis in Equation (7) can be estimated using the Laguerre Gaussian quadrature formula, which replaces the integral by a summation of terms over a certain number (say *K*) of support points, each term comprising the evaluation of the function $G_{qi}(.)$ at the support point *k* multiplied by a probability mass or weight associated with the support point [the support points are the roots of the Laguerre polynomial of order *K*, and the weights are computed based on a set of theorems provided by Press *et al.* (1992)].

3 THE MIXED MULTINOMIAL LOGIT (MMNL) CLASS OF MODELS

The HEV model in the previous section and the GEV models in Chapter 13 have the advantage that they are easy to estimate; the likelihood function for these models either includes a one-dimensional integral (in the HEV model) or is in closed-form (in the GEV models). However, these models are restrictive since they only partially relax the IID error assumption across alternatives. In this section, we discuss the MMNL class of models that are flexible enough to completely relax the independence and identically distributed error structure of the MNL as well as to relax the assumption of response homogeneity.

The mixed MMNL class of models involves the integration of the MNL formula over the distribution of unobserved random parameters. It takes the structure

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) \, f(\beta \mid \theta) \, d(\beta), \text{ where} \tag{11}$$

$$L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_{j} e^{\beta' x_{qi}}} .$$

P_{qi} is the probability that individual q chooses alternative i, x_{qi} is a vector of observed variables specific to individual q and alternative i, β represents parameters which are random realizations from a density function *f*(.), and θ is a vector of underlying moment parameters characterizing *f*(.).

The first applications of the mixed logit structure of Equation (11) appear to have been by Boyd and Mellman (1980) and Cardell and Dunbar (1980). However, these were not individual-level models and, consequently, the integration inherent in the mixed logit formulation had to be evaluated only once for the entire market. Train (1986) and Ben-Akiva *et al.* (1993) applied the mixed logit to customer-level data, but considered only one or two random coefficients in their specifications. Thus, they were able to use quadrature techniques for estimation. The first applications to realize the full potential of mixed logit by allowing several random coefficients simultaneously include Revelt and Train (1998) and Bhat (1998a), both of which were originally completed in early 1996 and exploited the advances in simulation methods.

The MMNL model structure of Equation (11) can be motivated from two very different (but formally equivalent) perspectives. Specifically, a MMNL structure may be generated from an intrinsic motivation to allow flexible substitution patterns across alternatives (error-components structure) or from a need to accommodate unobserved heterogeneity across individuals in their sensitivity to observed exogenous variables (random-coefficients structure).

3.1. Error-components Structure

The error-components structure partitions the overall random term associated with the utility of each alternative into two components: one that allows the unobserved error terms to be non-identical and non-independent across alternatives, and another that is specified to be independent and identically (type I extreme value) distributed across alternatives. Specifically, consider the following utility function for individual q and alternative i:

$$\begin{aligned} U_{qi} &= \gamma' y_{qi} + \zeta_{qi} \\ &= \gamma' y_{qi} + \mu' z_{qi} + \varepsilon_{qi} \end{aligned} \tag{12}$$

where $\gamma' y_{qi}$ and ζ_{qi} are the systematic and random components of utility, and ζ_i is further partitioned into two components, $\mu' z_{qi}$ and ε_{qi}. z_{qi} is a vector of observed data associated with alternative *i*, some of the elements of which might also appear in the vector y_{qi}. μ is a random vector with zero mean. The component $\mu' z_{qi}$ induces heteroscedasticity and correlation across unobserved utility components of the alternatives. Defining $\beta = (\gamma', \mu')'$ and $x_{qi} = (y'_{qi}, z'_{qi})'$, we obtain the MMNL model structure for the choice probability of alternative *i* for individual *q*.

The emphasis in the error-components structure is on allowing a flexible substitution pattern among alternatives in a parsimonious fashion. This is achieved by the “clever” specification of the variable vector z_{qi} combined with (usually) the specification of independent normally distributed random elements in the vector μ. For example, z_i may be specified to be a row vector of dimension M, with each row representing a group *m* (*m* = 1, 2, …, *M*) of alternatives sharing common unobserved components. The row(s) corresponding to the group(s) of which *i* is a member take(s) a value of one and other rows take a value of zero. The vector μ (of dimension *M*) may be specified to have independent elements, each element having a variance component σ_m^2. The result of this specification is a covariance of σ_m^2 among alternatives in group *m* and heteroscedasticity across the groups of alternatives. This structure is less restrictive than the nested logit structure in that an alternative can belong to more than one group. Also, by structure, the variance of the alternatives is different. More general structures for $\mu' z_i$ in equation (12) are presented by Ben-Akiva and Bolduc (1996) and Brownstone and Train (1999). Examples of the error-components motivation in the literature include Bhat (1998b), Jong *et al.* (2002a,b), Whelan *et al.* (2002), and Batley *et al.* (2001a,b). The reader is also referred to the work of Walker and her colleagues (Ben-Akiva *et al.*, 2001; Walker, 2002) and Munizaga and Alvarez-Daziano (2002) for important identification issues in the context of the error components MMNL model.

3.2. Random-coefficients Structure

The random-coefficients structure allows heterogeneity in the sensitivity of individuals to exogenous attributes. The utility that an individual *q* associates with alternative *i* is written as

$$U_{qi} = \beta_q' x_{qi} + \varepsilon_{qi} \tag{13}$$

where x_{qi} is a vector of exogenous attributes, β_q is a vector of coefficients that varies across individuals with density $f(\beta)$, and ε_{qi} is assumed to be an independently and identically distributed (across alternatives) type I extreme value error term. With this specification, the unconditional choice probability of alternative *i* for individual *q* is given by the mixed logit formula of equation (11). While several density functions may be used for *f*(.), the most commonly used is the normal distribution. A log-normal distribution may also be used if, from a theoretical perspective, an element of β has to take the same sign for every individual (such as a negative coefficient for the travel-time parameter in a travel-mode-choice model). Other distributions that have been used in the literature include triangular and uniform distributions (see Revelt and Train, 2000; Train, 2001; Hensher and Greene, 2003; Amador *et al.* 2005), the Rayleigh distribution (Siikamaki and Layton, 2001), the censored normal (Cirillo and Axhausen, 2006; Train and Sonnier, 2004), and Johnson's S_B (Cirillo and Axhausen, 2006; Train and Sonnier, 2004). The triangular and uniform distributions have the nice property that they are bounded on both sides, thus precluding the possibility of very high positive or negative coefficients for some decision-makers as would be the case if normal or log-normal distributions are used. By constraining the mean and spread to be the same, the triangular and uniform distributions can also be customized to cases where all decision-makers should have the same sign for one or more coefficients. The Rayleigh distribution, like the lognormal distribution, assures the same sign of coefficients for all decision-makers. The censored normal distribution is censored from below at a value, with a probability mass at that value and a density identical to the normal density beyond that value. This distribution is useful to simultaneously capture the influence of attributes that do not affect some individuals (*i.e.*, the individuals are indifferent) and affect other individuals. Johnson's S_B distribution is similar to the log-normal distribution, but is bounded from above and has thinner tails. Johnson's S_B can replicate a variety of distributions, making it a very flexible distribution. Its density can be symmetrical or asymmetrical, have a tail to the right or left, or become a flat plateau or be bi-modal[1].

The reader will note that the error-components specification in Equation (12) and the random-coefficients specification in Equation (13) are structurally equivalent. Specifically, if β_q is distributed with a mean of γ and deviation μ, then Equation (13) is identical to Equation (12) with $x_{qi} = y_{qi} = z_{qi}$. However, this apparent restriction for equality of Equations (12) and (13) is purely notational. Elements of x_{qi} that do not appear in z_{qi} can be viewed as variables the coefficients of which are deterministic in the population, while elements of x_{qi} that do not enter in y_{qi} may be viewed as variables the coefficients of which are randomly distributed in the population with mean zero.

3.3. Probability Expressions and General Comments

As indicated above, error-components and random-coefficients formulations are equivalent. Also, the random-coefficients formulation is more compact. Thus, we will adopt the random-coefficients notation to write the MMNL probability expression. Specifically, consider equation (13) and separate out the effect of

[1] The reader is referred to Hess and Axhausen (2005), Hess, Bielaire, and Polak (2005), and Train and Sonnier (2004) for a review of alternative distributional forms and their ability to approximate several different types of true distributional. Also, Sorenson and Nielson (2003) propose a method for determining the best distributional form prior to estimation.

variables with fixed coefficients (including the alternative specific constant) from the effect of variables with random coefficients, and write the utility function as:

$$U_{qi} = \alpha_{qi} + \sum_{k=1}^{K} \beta_{qk} x_{qik} + \varepsilon_{qi}, \tag{14}$$

where α_{qi} is the effect of variables with fixed coefficients. Let $\beta_{qk} \sim N(\mu_k, \sigma_k)$, so that $\beta_{qk} = \mu_k + \sigma_k s_{qk}$ (q = 1, 2, …, Q; k = 1, 2, …, K). In this notation, we are implicitly assuming that the β_{qk} terms are independent of one another. Even if they are not, a simple Choleski decomposition can be undertaken so that the resulting integration involves independent normal variates (see Revelt and Train, 1998). s_{qk} (q = 1, 2, …, Q; k = 1, 2, …, K) is a standard normal variate. Further, let $V_{qi} = \alpha_{qi} + \sum_k \mu_k x_{qik}$. The probability that the qth individual chooses alternative i for the random-coefficients logit model may be written as

$$P_{iq} = \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi}} + \sum_k \sigma_k s_{qk} x_{qik}}{\sum_j e^{V_{qj}} + \sum_k \sigma_k s_{qk} x_{qjk}} d\Phi(s_{q1}) d\Phi(s_{q2}) ... d\Phi(s_{qK}) \right\}, \tag{15}$$

where $\Phi(.)$ represents the standard normal cumulative distribution function and

The MMNL class of models can approximate any discrete choice model derived from random utility maximization (including the multinomial probit) as closely as one pleases (see McFadden and Train, 2000). The MMNL model structure is also conceptually appealing and easy to understand since it is the familiar MNL model mixed with the multivariate distribution (generally multivariate normal) of the random parameters (see Hensher and Greene, 2003). In the context of relaxing the IID error structure of the MNL, the MMNL model represents a computationally efficient structure when the number of error components (or factors) needed to generate the desired error covariance structure across alternatives is much smaller than the number of alternatives (see Bhat, 2003). The MMNL model structure also serves as a comprehensive framework for relaxing both the IID error structure as well as the response homogeneity assumption.

A few notes are in order here about the MMNL model vis-à-vis the MNP model. First, both these models are very flexible in the sense of being able to capture random taste variations and flexible substitution patterns. Second, both these models are able to capture temporal correlation over time, as would normally be the case with panel data. Third, the MMNL model is able to accommodate non-normal distributions for random coefficients, while the MNP model can handle only normal distributions. Fourth, researchers and practitioners familiar with the traditional MNL model might find it conceptually easier to understand the structure of the MMNL model compared to the MNP. Fifth, both the MMNL and MNP model, in general, require the use of simulators to estimate the multidimensional integrals in the likelihood function. Sixth, the MMNL model can be viewed as arising from the use of a logit-smoothed Accept-Reject (AR) simulator for an MNP model (see Bhat 2000, and Train 2003; page 124). Seventh, the simulation techniques for the MMNL model are conceptually simple, and straightforward to code. They involve simultaneous draws from the appropriate density function with unrestricted ranges for all alternatives. Overall, the MMNL model is very appealing and broad in scope, and there appears to be little reason to prefer the MNP model over the MMNL model. However, there is at least one exception to this general rule, corresponding to the case of normally distributed random taste coefficients. Specifically, if the number of normally distributed random coefficients is substantially more than the number of alternatives, the MNP model offers advantages because the dimensionality is of the order of the number of alternatives (in the MMNL, the dimensionality is of the order of the number of random coefficients)[2].

[2] The reader is also referred to Munizaga and Alvarez-Daziano (2002) for a detailed discussion comparing the MMNL model with the nested logit and MNP models.

4 THE MIXED GEV CLASS OF MODELS

The MMNL class of models is very general in structure and can accommodate both relaxations of the IID assumption as well as unobserved response homogeneity within a simple unifying framework. Consequently, the need to consider a mixed GEV class may appear unnecessary. However, there are instances when substantial computational efficiency gains may be achieved using a MGEV structure that superimposes a mixing distribution over an underlying GEV model rather than over the MNL model. Consider, for instance, Bhat and Guo's (2004) model for household residential location choice. It is possible, if not very likely, that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow contiguous alternatives to be correlated. In the MMNL structure, such a correlation structure may be imposed through the specification of a multivariate MNP-like error structure, which will then require multidimensional integration of the order of the number of spatial units (see Bolduc *et al.*, 1996). On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation[3]. However, the GEV model structure of Bhat and Guo cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such random coefficients, leading to a parsimonious and powerful MGEV structure. Thus, in a case with 1000 spatial units (or zones), the MMNL model would entail a multidimensional integration of the order of 1000 plus the number of random coefficients, while the MGEV model involves multidimensional integration only of the order of the number of random coefficients (a reduction of dimensionality of the order of 1000!).

[3] The GEV structure used by Bhat and Guo is a restricted version of the GNL model proposed by Wen and Koppelman (2001). Specifically, the GEV structure takes the form of a paired GNL (PGNL) model with equal dissimilarity parameters across all paired nests (each paired nest includes a spatial unit and one of its adjacent spatial units).

In addition to computational efficiency gains, there is another more basic reason to prefer the MGEV class of models when possible over the MMNL class of models. This is related to the fact that closed-form analytic structures should be used whenever feasible, because they are always more accurate than the simulation evaluation of analytically intractable structures (see Train, 2003; pg. 191). In this regard, superimposing a mixing structure to accommodate random coefficients over a closed form analytic structure that accommodates a particular desired inter-alternative error correlation structure represents a powerful approach to capture random taste variations and complex substitution patterns.

Clearly, there are valuable gains to be achieved by combining the state-of-the-art developments in closed-form GEV models with the state-of-the-art developments in open-form mixed distribution models. With the recent advances in simulation techniques, there appears to be a feeling among some discrete choice modelers that there is no need for any further consideration of closed-form structures for capturing correlation patterns. But, as Bhat and Guo (2004) have demonstrated in their paper, the developments in GEV-based structures and open-form mixed models are not as mutually exclusive as may be the impression in the field; rather these developments can, and are, synergistic, enabling the estimation of model structures that cannot be estimated using GEV structures alone or cannot be efficiently estimated (from a computational standpoint) using a mixed multinomial logit structure.

5 SIMULATION ESTIMATION TECHNIQUES

The mixed models discussed in Sections 3 and 4 require the evaluation of analytically intractable multidimensional integrals in the classical estimation approach. The approximation of these integrals is undertaken using simulation techniques that entail the evaluation of the integrand at a number of draws taken from the domain of integration (usually the multivariate normal distribution) and computing the average of the resulting integrand values across the different draws. The draws can be taken by generating standard univariate draws for each dimension, and developing the necessary multivariate draws through a simple Cholesky decomposition of the target multivariate covariance matrix applied to the standard univariate draws. Thus, the focus of simulation techniques is on generating N sets of S univariate draws for each individual, where N is the number of draws and S is the dimensionality of integration. To maintain independence over the simulated likelihood functions of decision-makers, different draws are used for each individual.

Three broad simulation methods are available for generating the draws needed for mixed model estimations: (a) Monte Carlo methods, (b) Quasi-Monte Carlo methods, and (c) Randomized Quasi-Monte Carlo methods. Each of these is discussed descriptively below. Mathematical details are available in Bhat (2001; 2003), Sivakumar *et al.* (2005), and Train (2003; Chapter 9).

5.1. The Monte-Carlo Method

The Monte-Carlo simulation method (or “the method of statistical trials”) to evaluating multidimensional integrals entails computing the integrand at a sequence of “random” points and computing the average of the integrand values. The basic principle is to replace a continuous average by a discrete average over randomly chosen points. Of course, in actual implementation, truly random sequences are not available; instead, deterministic pseudo-random sequences which appear random when subjected to simple statistical tests are used (see Niederreiter, 1995 for a discussion of pseudo-random sequence generation). This pseudo-Monte Carlo (or PMC) method has a slow asymptotic convergence rate with the expected integration error of the order of $N^{-0.5}$ in probability (N being the number of pseudo-random points drawn from the s-dimensional integration space). Thus, to obtain an added decimal digit of accuracy, the number of draws needs to be increased hundred fold. However, the PMC method's convergence rate is remarkable in that it is applicable for a wide class of integrands (the only requirement is that the integrand have a finite variance; see Spanier and Maize, 1991). Further, the integration error can be easily estimated using the sample values and invoking the central limit theorem, or by replicating the evaluation of the integral several times using independent sets of PMC draws and computing the variance in the different estimates of the integrand.

5.2. The Quasi-Monte Carlo Method

The quasi-Monte Carlo method is similar to the Monte Carlo method in that it evaluates a multidimensional integral by replacing it with an average of values of the integrand computed at discrete points. However, rather than using pseudo-random sequences for the discrete points, the quasi-Monte Carlo approach uses "cleverly" crafted non-random and more uniformly distributed sequences (labeled as quasi-Monte Carlo or QMC sequences) within the domain of integration. The underlying idea of the method is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space. The convergence rate for quasi-random sequences is, in general, faster than for pseudo-random sequences. In particular, the theoretical upper bound of the integration error for reasonably well-behaved smooth functions is of the order of N^{-1} in the QMC method, where N is the number of quasi-random integration points.

The QMC sequences have been well known for a long time in the number theory literature. However, the focus in number theory is on the use of QMC sequences for accurate evaluation of a single multidimensional integral. In contrast, the focus of the maximum simulated likelihood estimation of econometric models is on accurately estimating underlying model parameters through the evaluation of multiple multidimensional integrals, each of which involves a parameterization of the model parameters and the data. The intent in the latter case is to estimate the model parameters accurately, and not expressly on evaluating each integral itself accurately.

Bhat (2001) proposed and introduced, in 1999, a simulation approach using QMC sequences for estimating discrete choice models with analytically intractable likelihood functions. There are several quasi-random sequences that may be employed in the QMC simulation method. Among these sequences are those that belong to the family of r-*adic* expansion of integers: the Halton, Faure, and Sobol sequences (see Bratley *et al*., 1992 for a good review). Bhat used the Halton sequence in the QMC simulation because of its conceptual simplicity. In his approach, Bhat generates a multidimensional QMC sequence of length $N*Q$, then uses the first N points to compute the contribution of the first observation to the criterion function, the second N points to compute the contribution of the second observation, and so on. This technique is based on averaging out of simulation errors across observations. But rather than being random sets of points across observations, each set of N points fills in the gaps left by the sets of N points used for previous observations. Consequently, the averaging effect across observations is stronger when using QMC sequences than when using the PMC sequence. In addition to the stronger averaging out effect across observations, the QMC sequence also provides more uniform coverage over the domain of the integration space for each observation compared to the PMC sequence. This enables more accurate computations of the probabilities for each observation with fewer points (*i.e.*, smaller N) when QMC sequences are used.

Bhat compared the Halton and PMC sequences in their ability to accurately and reliably recover model parameters in a mixed logit model. His experimental and computational results indicated that the Halton sequence outperformed the PMC sequence by a substantial margin. Specifically, he found that 125 Halton draws produced more accurate parameters than 2000 PMC draws in estimation, and noted that this substantial reduction in computational burden can dramatically influence the use of mixed models in practice. Subsequent studies by Train (2000), Hensher (2001a), Munizaga and Alvarez-Daziano (2001), and Jong *et al.* (2002a,b) have confirmed this dramatic improvement using the Halton sequence. For example, Hensher (2001a) found that the data fit and parameter values of the mixed logit model in his study remained about the same beyond 50 Halton draws and concludes that the QMC approach is “a phenomenal development in the estimation of complex choice models”.

Sandor and Train (2004) have found that there is some room for further improvement in accuracy and efficiency using more complex digital QMC sequences proposed by Niederreiter and his colleagues relative to the Halton sequence. Bhat (2003) suggests a scrambled Halton approach in high dimensions to reduce the correlation along high dimensions of a standard Halton sequence (see also Braaten and Weller, 1979), and shows that the scrambling improves the performance of the standard Halton sequence.

A limitation of the QMC method for simulation estimation, however, is that there is no straightforward practical way of statistically estimating the error in integration, because of the deterministic nature of the QMC sequences. Theoretical results are available to compute the upper bound of the error using a well-known theorem in number theory referred to as the Koksma-Hlawka inequality (Zaremba, 1968). But, computing this theoretical error bound is not practical and, in fact, is much more complicated than evaluating the integral itself (Owen, 1997; Tuffin, 1996). Besides, the upper bound of the integration error from the theoretical result can be very conservative (Owen, 1998).

5.3. The Hybrid Method

The discussion in the previous two sections indicates that QMC sequences provide better accuracy than PMC sequences, while PMC sequences provide the ability to estimate the integration error easily. To take advantage of the strengths of each of these two methods, it is desirable to develop hybrid or randomized QMC sequences (see Owen, 1995 for a history of such hybrid sequences). The essential idea is to introduce some randomness into a QMC sequence, while preserving the equidistribution property of the underlying QMC sequence. Then, by using several independent randomized QMC sequences, one can use standard statistical methods to estimate integration error.

Bhat (2003) describes a process to randomize QMC sequences for use in simulation estimation. This process, based on Tuffin's (1996) randomization procedures, is described intuitively and mathematically by Bhat in the context of a single multidimensional integral. Sivakumar *et al.* (2005) experimentally compared the performance of revised hybrid sequences based on the Halton and Faure sequences in the context of the simulated likelihood estimation of an MMNL model of choice. They also assessed the effects of scrambling on the accuracy and efficiency of these sequences. In addition, they compared the efficiency of the QMC sequences generated with and without scrambling across observations. The results of their analysis indicate that the Faure sequence consistently outperforms the Halton sequence. The Random Linear and Random Digit scrambled Faure sequences, in particular, are among the most effective QMC sequences for simulated maximum likelihood estimation of the MMNL model.

5.4. Summary on Simulation Estimation of Mixed Models

The discussion above shows the substantial progress in simulation methods, and the arrival of quasi-Monte Carlo (QMC) methods as an important breakthrough in the simulation estimation of advanced discrete choice models. The discovery and application of QMC sequences for discrete choice model estimation is a watershed event and has fundamentally changed the way we think about, specify, and estimate discrete choice models. In the very few years since it was proposed by Bhat at the turn of the millennium, it has already become the "bread and butter" of simulation techniques in the field.

6 CONCLUSIONS AND APPLICATION OF ADVANCED MODELS

This chapter has discussed the structure, estimation techniques, and transport applications of three different classes of discrete choice models — heteroscedastic models, mixed multinomial logit (MMNL) models, and mixed generalized extreme value models. The formulations presented are quite flexible although estimation using the maximum likelihood technique requires the evaluation of one-dimensional integrals (in the HEV model) or multi-dimensional integrals (in the MMNL and MGEV models). However, these integrals can be approximated using Gaussian quadrature techniques or simulation techniques. The advent of fast computers and the development of increasingly more efficient sequences for simulation have now made the estimation of such analytically intractable model formulations very practical. In this regard, QMC simulation techniques have proved to be very effective. This should be evident from Table 1, which lists recent (within the past 5 years) transportation applications of flexible discrete choice models. There is a clear shift from pseudo-random draws to QMC draws (primarily Halton draws) in the more recent applications of flexible choice structures. Additionally, Table 1 illustrates the wide applicability of flexible choice structures, including airport operations and planning, travel behavioral analysis, travel mode choice, and other transport-related fields.

A note of caution before closing. It is important for the analyst to continue to think carefully about model specification issues rather than to use the (relatively) advanced model formulations presented in this chapter as a panacea for all systematic specification ills. The flexible models presented here should be viewed as formulations that recognize the inevitable presence of unobserved heterogeneity in individual responsiveness across individuals and/or of interactions among unobserved components affecting the utility of alternatives (because it is impossible to identify, or collect data on, all factors affecting choice decisions). The flexible models are not, however, a substitute for careful identification of systematic variations in the population. The analyst must always explore alternative and improved ways to incorporate systematic effects in a model. The flexible structures can then be superimposed on models that have attributed as much heterogeneity to systematic variations as possible. Another important issue in using flexible models is that the specification adopted should be easy to interpret; the analyst would do well to retain as simple a specification as possible while attempting to capture the salient interaction patterns in the empirical context under study. The MMNL model is particularly appealing in this regard since it “forces” the analyst to think structurally during model specification.

The confluence of continued careful structural specification with the ability to accommodate very flexible substitution patterns or unobserved heterogeneity should facilitate the application of behaviorally rich structures in transportation-related discrete choice modeling in the years to come.

References

Adler, T., C.S. Falzarano, Spitz, G. (2005). Modeling service trade-offs in air itinerary choices, *Transportation Research Record*, 1915.

Amador, F.J., R.M. Gonzalez, and Ortuzar, J.D. (2005). Preference heterogeneity and willingness to pay for travel time savings, *Transportation*, 32, 627-647.

Batley, R., Fowkes, T., Watling, D., Whelan, G., Daly, A. and Hato, E. (2001a). Models for analysing route choice, Paper presented at the 33rd Annual Conference of the Universities Transport Studies Group Conference, University of Oxford.

Batley, R., Fowkes, T., Whelan ,G. and Daly, A. (2001b). Models for choice of departure time, Paper presented to the European Transport Conference, Association of European Transport, University of Cambridge.

Bajwa, S., Bekhor, S., Kuwahara, M. and E. Chung (2006). Discrete choice modeling of combined mode and departure time, paper presented at the 11th International Conference on Travel Behavior Research, Kyoto, August 2006.

Bekhor, S., Ben-Akiva, M. and Ramming, M.S. (2002), Adaptation of logit kernel to route choice situation, *Transportation Research Record*, 1805, 78-85.

Ben-Akiva, M. and Bolduc, D. (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, Department of Civil Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, and Départmente d'Economique, Université Laval, Sainte-Foy, QC, working paper.

Ben-Akiva, M., Bolduc, D. and M. Bradley (1993). Estimation of travel model choice models with randomly distributed values of time, *Transportation Research Record*, 1413, 88-97.

Ben-Akiva, M., Bolduc, D., and Walker, J. (2001). Specification, estimation and identification of the logit kernel (or continuous mixed logit) model, Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

Bhat, C.R. (1995). A heteroscedastic extreme-value model of intercity mode choice, *Transportation Research B*, 29(6), 471-483.

Bhat, C.R. (1998a). Accommodating variations in responsiveness to level-of-service variables in travel mode choice modeling, *Transportation Research Part A*, 32 (7), 495-507.

Bhat, C.R. (1998b). Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel mode and departure time choice, *Transportation Research Part B*, 32 (7), 455-466.

Bhat, C.R. (2000). A multi-level cross-classified model for discrete response variables, *Transportation Research Part B*, 34 (7), 567-582.

Bhat, C.R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, *Transportation Research Part B*, 35 (7), 677-693.

Bhat, C.R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences, *Transportation Research Part B*, 37 (9), 837-855.

Bhat, C.R. (2007). Econometric choice formulations: Alternative model structures, estimation techniques, and emerging directions, *Moving Through Nets: The Physical and Social Dimensions of Travel - Selected papers from the 10th International Conference on Travel Behaviour Research*, Chapter 3, pp. 45-80, Edited by K. W. Axhausen, Elsevier.

Bhat, C.R. and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco Bay area, *Transportation Research Part B*, 36 (7), 577–669.

Bhat, C.R. and Gossen, R. (2004). A mixed multinomial logit model analysis of weekend recreational episode type choice, *Transportation Research Part B*, 38 (9), 767-787.

Bhat, C.R. and Guo, J. (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling, *Transportation Research Part B*, 38 (2), 147-168.

Bhat, C.R. and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice, *Transportation Research Part B*, 40, 709-730.

Bhat, C.R. and Zhao, H. (2002). The spatial analysis of activity stop generation, *Transportation Research Part B*, 36 (6), 557-575.

Bolduc, D., Fortin, B. and Fournier, M. (1996). The effect of incentive policies on the practice location of doctors: A multinomial probit analysis, *Journal of Labor Economics*, 14 (4), 703-732.

Boyd, J. and Mellman, J. (1980). The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis, *Transportation Research Part A*, 14, 367-378.

Braaten, E. and Weller G. (1979). An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration, *Journal of Computational Physics*, 33 (2), 249-258.

Bratley, P. Fox, B.L., and Niederreiter, H. (1992). Implementation and tests of low-discrepancy sequences, *ACM Transactions on Modeling and Computer Simulation*, 2 (3), 195-213.

Brownstone, D. and Small, K.A. (2005). Valuing time and reliability: Assessing the evidence from road pricing demonstrations, *Transportation Research Part A*, 39 (4), 279-293.

Brownstone, D. and Train, K. (1999). Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, 89, 109-129.

Cardell, S. and Dunbar, F. (1980). Measuring the societal impacts of automobile downsizing, *Transportation Research Part A*, 14, 423-434.

Carlsson, F. (2003). The demand for intercity public transport: the case of business passengers, *Applied Economics*, 35, 41-50.

Chamberlain, G. (1980). Analysis of covariance with qualitative data, *Review of Economic Studies*, 47, 225-238.

Cirillo, C. and Axhausen, K.W. (2006). Evidence on the distribution of values of travel time savings from a six-week diary, *Transportation Research Part A*, 40, 444-457.

Daganzo, C. (1979). *Multinomial probit: The theory and its application to demand forecasting*. New York: Academic Press.

Galilea, P. and Ortúzar, J. D. (2005). Valuing noise level reductions in a residential location context, *Transportation Research Part D*, 4, 305-322.

Greene, W.H., Hensher, D.A. and Rose, J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models, *Transportation Research Part B*, 40, 75-92.

Han, B., Algers, S., and Engelson, L. (2001). Accommodating drivers' taste variation and repeated choice correlation in route choice modeling by using the mixed logit model, Presented at the 80th Annual Meeting of the Transportation Research Board, Washington D.C.

Hensher, D.A. (2001a). Measurement of the valuation of travel time savings, *Journal of Transport Economics and Policy*, 35 (1), 71-98.

Hensher (2001b). The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications, *Transportation*, 28 (2), 101–118.

Hensher, D.A. (2006). Towards a practical method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions, *Transportation Research Part A,* 40, 829-840.

Hensher, D.A. and Greene, W. (2003). The mixed logit model: The state of practice. *Transportation*, 30 (2), 133-176.

Hess, S. and Axhausen, K.W. (2005). Distributional assumptions in the representation of random taste heterogeneity, presented at the 5th Swiss Transport Research Conference (STRC), Monte Verita, March 2005.

Hess, S., Bierlaire, M. and Polak, J.W. (2004). Development and application of a mixed cross-nested logit model, Proceedings of the XXIth European Transport Conference.

Hess, S., Bierlaire, M., and Polak, J.W. (2005). Estimation of value of travel-time savings using mixed logit models, *Transportation Research Part A,* 39, 221-236.

Hess, S. and Polak, J.W. (2005). Mixed logit modeling of airport choice in multi-airport regions, *Journal of Air Transportation Management*, 11 (2), 59-68.

Iragüen, P. and Ortúzar, J.D. (2004). Willingness-to-pay for reducing fatal accident risk in urban areas: An internet-based web page stated preference survey, *Accident Analysis and Prevention*, 36 (4), 513-524.

Jong, G. de, Pieters, M. and Daly, A. (2002a). Testing models for time of day choice with individual-specific effects, Working Paper, RAND Europe, Leiden.

Jong, G. de, Daly, A., Pieters, M., Vellay, C., Bradley, M. and Hofman, F. (2002b). A model for time of day and mode choice using error components logit, Working Paper, RAND Europe, Leiden.

Lapparent, M. (2006). The choice of a mode of transportation for home to work trips in the French Parisian region: application of mixed GEV models within non linear utility functions, paper presented at the 11th International Conference on Travel Behavior Research, Kyoto, August 2006.

Lee, J.H., Chon, K.S. and Park, C. (2004). Accommodating heterogeneity and heteroscedasticity in intercity travel mod choice model: formulation and application to Honam, South Korea, high speed rail demand analysis, Transportation Research Record, 1898.

Lijesen, M.G. (2006). A mixed logit based valuation of frequency in civil aviation from SP-data, *Transportation Research Part E*, 42 (2), 82-94.

Luce, R. and Suppes, P. (1965). Preference, utility and subjective probability, in: R. Luce, R. Bush and E. Galanter, eds, *Handbook of mathematical psychology*, Vol. 3. New York: Wiley.

McFadden, D. and Train, K. (2000). Mixed MNL models of discrete response, *Journal of Applied Econometrics*, 15 (5), 447-470.

Mohammadian, H. and Doherty, S.T. (2005). Mixed logit model of activity-scheduling time horizon incorporating spatial-temporal flexibility variables, *Transportation Research Record*, 1926, 33-40.

Munizaga, M. and Alvarez-Daziano, R. (2001). Mixed logit versus nested logit and probit, Working Paper, Departmento de Ingeniera Civil, Universidad de Chile.

Munizaga, M. and Alvarez-Daziano, R. (2002). Evaluation of mixed logit as a practical modelling alternative, Proceedings European Transport Conference, Cambridge, UK.

Niederreiter, H. (1995). New developments in uniform pseudo-random number and vector generation, in H. Niederreiter and J-S. Shiue (eds.) *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computin*g, 94-112, Springer, New York.

Owen, A.B. (1995). Randomly permuted (t,m,s)-nets and (t,s)-sequences, in H. Niederreiter and J-S. Shiue (eds.) *Monte Carlo Methods in Scientific Computin*g, 299-317, Springer, New York.

Owen, A.B (1997). Scrambled net variance for integrals of smooth functions, *The Annals of Statistic*s, 25 (4), 1541-562.

Owen, A.B. (1998). Latin supercube sampling for very high dimensional simulations, *ACM Transactions on Modeling and Computer Simulatio*n, 8, 71-102.

Pathomsiri, S. and Haghani, A. (2005). Taste variations in airport choice models, *Transportation Research Record*, 1915.

Pinjari, A.R., and Bhat, C.R. (2006). On the nonlinearity of response to level of service variables in travel mode choice models, forthcoming, *Transportation Research Record.*

Press, W.H., Teukolsky, S.A. and Nerlove, M. (1992). *Numerical recipes in C: The art of scientific computing*. Cambridge, MA: Cambridge University Press.

Recker, W.W. (1995). Discrete choice with an oddball alternative, *Transportation Research B*, 29, 201-212.

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

Revelt, D. and Train, K. (2000). Customer-specific taste parameters and mixed logit, Working Paper No. E00-274, Department of Economics, University of California, Berkeley.

Rizzi, L.I. and Ortúzar, J. D. (2003). Stated preference in the valuation of interurban road safety, *Accident Analysis and Prevention*, 35 (1), 9-22.

Siikamaki, J. and Layton, D. (2001). Pooled models for contingent valuation and contingent ranking data: valuing benefits from biodiversity conservation, Working Paper, Department of Agricultural and Resource Economics, University of California, Davis.

Silliano, M. and Ortúzar, J. D. (2005). Willingness-to-pay estimation with mixed logit models: Some new evidence, *Environment and Planning A*, 37 (3) 525-550.

Sivakumar, A., and Bhat, C.R. (2006). A comprehensive, unified, framework for analyzing spatial location choice, Technical paper, Department of Civil Engineering, The University of Texas at Austin, August 2006.

Sivakumar, A., Bhat, C.R. and Ökten, G. (2005). Simulation estimation of mixed discrete choice models with the use of randomized quasi-monte carlo sequences: a comparative study, *Transportation Research Record*, 1921, 112-122.

Small, K., Winston, C. and Yan, J. (2005). Uncovering the distribution of motorists’ preferences for travel time and reliability: Implications for road pricing, *Econometrica*, 73 (4), 1367-1382.

Sorensen, M.V. and Nielson, O.A. (2003). MSL for mixed logit model estimation – on shape of distributions, *In Proceedings of European Transport Conference*, Strasbourg, France.

Spanier, J. and Maize, E. (1991). Quasi-random methods for estimating integrals using relatively small samples, *SIAM Review*, 36 (1), 18-44.

Srinivasan, K.K. and Athuru, S.R. (2005). Analysis of within-household effects and between-household differences in maintenance activity allocation, *Transportation*, 32, 495-521.

Srinivasan, K.K. and Mahmassani, H.S. (2003). Analyzing heterogeneity and unobserved structural effects in route-switching behavior under ATIS: A dynamic kernel logit formulation, *Transportation Research Part B*, 37 (9), 793-814.

Srinivasan, K.K. and Ramadurai, G. (2006). Dynamics and variability in within-day mode choice decisions: role of state dependence, habit persistence, and unobserved heterogeneity, Presented at the Transportation Research Board 85th Annual Meeting, Washington D.C.

Train, K. (1986). *Qualitative Choice Analysis*, MIT Press, Cambridge.

Train, K. (2000). Halton sequences for mixed logit, Working Paper E00-278, Department of Economics, University of California, Berkeley.

Train, K. (2001). A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit, Working Paper, Department of Economics, University of California, Berkeley.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Tuffin, B. (1996). On the use of low discrepancy sequences in Monte Carlo methods, *Monte Carlo Methods and Applications*, 2 (4), 295-320.

Vichiensan, V., Miyamoto, K. and Tokunaga, Y. (2005). Mixed logit model framework with structuralized spatial effects: a test of applicability with area unit systems in location analysis, *Journal of Eastern Asia Society for Transportation Studies*, 6, 3789-3802.

Walker, J.L. (2002). Mixed logit (or logit kernel) model: dispelling misconceptions of identification, *Transportation Research Record*, 1805, 86-98.

Walker, J.L. and Parker, R.G. (2006). Estimating utility of time-of-day for airline schedules using mixed logit model, Presented at the Transportation Research Board 85th Annual Meeting, Washington D.C.

Warburg, V., Bhat, C.R. and Adler, T. (2006). Modeling demographic and unobserved heterogeneity in air passengers' sensitivity to service attributes in itinerary choice, *Transportation Research Record*, 1951, 7-16.

Wen, C-H. and Koppelman, F.S. (2001). The generalized nested logit model, *Transportation Research Part B*, 35 (7), 627-641.

Whelan, G., Batley, R., Fowkes, T. and Daly, A. (2002). Flexible models for analyzing route and departure time choice, European Transport Conference Proceedings, Association for European Transport, Cambridge.

Zaremba, S.K. (1968). The mathematical basis of Monte Carlo and quasi-Monte Carlo methods, *SIAM Review*, 10 (3), 303-314.

Table 1. Sample of Recent (within the past 5 years) Travel Behavior Applications of Advanced Discrete Choice Models

Model Type	Authors	Model Structure	Application Focus	Data Source	Type of Simulation Draws
HEV	Hensher (2006)	Heteroscedastic error terms	Route choice: Accommodating scale differences of varying SP data designs through unconstrained variances on the random components of each alternative	2002 SP travel survey conducted in Sydney, Australia	--
MMNL	Bekhor *et al.* (2002)	Error components structure	Travel route choice: Accommodating unobserved correlation on paths with overlapping links.	1997 transportation survey of MIT faculty and staff.	Pseudo-random draws
	Jong *et al.* (2002a)	Error components structure	Travel mode and time-of-day choice: Allowing unobserved correlation across time and mode dimensions.	2001 SP data collected from travelers during extended peak periods (6-11 a.m. and 3-7 p.m.) on weekdays.	Pseudo-random draws
	Vichiensan, Miyamoto, and Tokunaga (2005)	Error components structure	Residential location choice: Accommodates spatial dependency between residential zones by specifying spatially autocorrelated deterministic and random error components.	2002 RP urban travel survey data collected in Sendai City, Japan.	Pseudo-random draws
	Amador, Gonzalez, and Ortuzar (2005)	Random coefficients structure	Mode choice: Accommodating unobserved individual-specific sensitivities to travel time and other factors.	2000 survey of economic and business students' mode choice to school collected in La Laguna, Spain.	Halton draws
	Bhat and Sardesai (2006)	Random coefficients structure	Commute mode choice: Accommodating scale differences between SP and RP choices and accounting for unobserved individual-specific sensitivities to travel time and reliability variables.	2000 RP/SP simulator-based experiment with Austin area commuters.	Halton draws
	Han *et al.* (2001)	Random coefficients structure	Travel route choice: Incorporating unobserved individual-specific heterogeneity to route choice determinants (delay, heavy traffic, normal travel time, *etc.*).	2000 SP survey and scenario data collected in Sweden.	Pseudo-random draws
	Hensher (2001a)	Random coefficients structure	Long distance travel route choice: Accommodating unobserved individual-specific sensitivities to different components of travel time (free flow time, slowed-down time, and stop time).	2000 SP survey data collected in New Zealand.	Pseudo-random draws

MMNL	Brownstone and Small (2005)	Random coefficients structure	Choice of toll versus non-toll facility: Allowing random coefficients to account for individual-specific unobserved preferences, and responsiveness to travel time and unreliability of travel time.	1996-2000 RP/SP survey from the SR-91 facility in Orange County, California.	Pseudo-random draws
	Carlsson (2003)	Random coefficients structure	Mode choice: Allowing coefficients to vary for each individual across choice situation and allowing for individual-specific unobserved preferences for specific modes and other factors.	SP intercity travel survey of business travelers between Stockholm and Gothenburg.	Pseudo-random draws
	Cirillo and Axhausen (2006)	Random coefficients structure	Mode choice: Accommodating unobserved individual-specific sensitivities to travel time and other factors and accounting for correlation across tours for the same individual.	1999 multi-week urban travel survey collected in Karlsruhe and Halle, Germany.	Halton draws
	Iragüen and Ortúzar (2004)	Random coefficients structure	Urban route choice: Recognizing unobserved individual heterogeneity in sensitivities to cost, number of accidents, and travel time.	2002 SP survey of car users of several private and public employment firms in Santiago.	Information not provided
	Galilea and Ortúzar (2005)	Random coefficients structure	Residential location choice: Accommodating unobserved individual heterogeneity in sensitivities to travel time to work, monthly rent, and noise level.	2002 SP survey of a sample of Santiago residents.	Information not provided
	Greene, Hensher, and Rose (2006)	Random coefficients structure	Commuter Mode Choice: Parameterizing the variance heterogeneity to examine the moments associated with the willingness to pay for travel time savings.	2003 SP survey of transport mode preferences collected in New South Wales, Australia.	Halton draws
	Hensher and Greene (2003)	Random coefficients structure	Urban commute travel route choice: Accommodating unobserved individual-specific sensitivities to different components of travel time and cost.	1999 SP survey data sets collected in seven cities in New Zealand.	Halton draws
	Hensher (2001b)	Random coefficients structure	The valuation of commuter travel time savings for car drivers: Comparing the value of travel savings obtained from MNL and alternative specifications of mixed logit models.	1999 SP/RP survey of residents in New Zealand.	Halton draws

MMNL	Hess *et al.* (2005)	Random coefficients structure	Travel time savings: Addressing the issue of non-zero probability of positive travel-time coefficients within the context of mixed logit specifications.	1989 Rail Operator data in the Toronto –Montreal corridor, Canada	Information not provided
	Hess and Polak (2005)	Random coefficients structure	Airport choice: Accommodating taste heterogeneity associated with the sensitivity to access time in choosing a departing airport.	1995 Airline passenger survey collected in the San Francisco Bay area.	Halton draws
	Lijesen (2006)	Random coefficients structure	Valuation of frequency in aviation: Developing a framework to link flight frequency with optimal arrival time and accounting for heterogeneity within customers' valuation of schedule delay.	Conjoint choice analysis experiment.	Information not provided
	Mohammadian and Doherty (2004)	Random coefficients structure	Choice of activity scheduling time horizon: Accommodating unobserved individual-specific sensitivities to travel time, flexibility in time, and activity frequency.	2002-2003 household activity scheduling survey collected in Toronto, Canada.	Pseudo-random draws
	Pathomsiri and Haghani (2005)	Random coefficients structure	Airport choice: Capturing random taste variations across passengers in response to airport level of service.	1998 Air passenger survey database for Baltimore, Washington DC	Information not provided
	Rizzi and Ortúzar (2003)	Random coefficients structure	Urban and interurban route choice: Accommodating unobserved individual heterogeneity in sensitivities to toll, travel time, and accidents.	2002 stated choice survey collected in Santiago and 1999-2000 survey collected in Santiago, Vina del Mar, Valparaiso, and Rancagua .	Information not provided
	Silliano and Ortúzar (2005)	Random coefficients structure	Residential choice incorporating unobserved individual heterogeneity in sensitivities to travel time to work, travel time to school, and days of alert status associated with the air quality of the zone of dwelling unit.	2001 SP survey conducted in Santiago.	Information not provided
	Small *et al.* (2005)	Random coefficients structure	Use of toll facilities versus non-toll facilities. Allowing random coefficients to accommodate unobserved individual-specific preferences and sensitivities to cost, travel time, and reliability.	1996-2000 RP/SP survey from the SR-91 facility in Orange County, California.	Pseudo-random draws

	Sivakumar and Bhat (2006)	Random coefficients structure	Spatial location choice: Developing a framework for modeling spatial location choice incorporating spatial cognition, heterogeneity in preference behavior, and spatial interaction.	1999 Travel survey in Karlsruhe (West Germany) and Halle (East Germany)	Random Linear scrambled Faure sequence
MMNL	Valdemar *et al.* (2005)	Random coefficients structure	Air passenger sensitivity to service attributes: Accommodating observed heterogeneity (related to demographic- and trip-related factors) and residual heterogeneity (related to unobserved factors).	2001 online survey of air travelers in US.	Halton draws
	Walker and Parker (2006)	Random coefficients structure	Time of day Airline demand: Formulating a continuous time utility function for airline demand.	2004 stated preference survey conducted by Boeing	Information not provided
	Adler *et al.* (2005)	Error components and random coefficients structure	Air itinerary choices: Modeling service tradeoffs by including the effects of itinerary choices of airline travel, airport, aircraft type and their corresponding interactions.	2000 Stated Preference survey of US domestic air travelers	Halton Draws
	Bhat and Castelar (2002)	Error components and random coefficients structure	Mode and time-of-day choice: Allowing unobserved correlation across alternatives through error components, preference heterogeneity and variations in responsiveness to level-of-service through random coefficients, and inertia effects of RP choice on SP choices through random coefficients.	1996 RP/SP multiday urban travel survey from the San Francisco Bay area.	Halton draws
	Bhat and Gossen (2004)	Error components and random coefficients structure	Weekend recreational episode type choice: Recognizing unobserved correlation in out-of-home episode type utilities and unobserved individual-specific preferences to participate in in-home, away-from-home, and recreational travel episodes.	2000 RP multiday urban travel survey collected in the San Francisco Bay area.	Halton draws
	Jong *et al.* (2002b)	Error components and random coefficients	Travel mode and time-of-day choice: Allowing unobserved correlation across time and mode dimensions; individual specific random effects.	2001 SP data collected from travelers during extended peak periods (6-11 a.m. and 3-7 p.m.) on weekdays.	Pseudo-random draws
	Lee *et al.* (2004)	Error components and random coefficients structure	Travel mode choice: Accommodating heterogeneity and heteroscedasticity in intercity travel mode choice.	RP/SP survey of users from Honam, South Korea.	Halton draws

	Pinjari and Bhat (2005)	Error components and random coefficients structure	Travel mode choice: Incorporating non-linearity of response to level of service variables for travel mode choice.	2000 RP/SP simulator-based experiment with Austin area commuters.	Halton draws
MMNL	Srinivasan and Mahmassani (2003)	Error components and random coefficients structure	Route switching behavior under Advanced Traveler Information System (ATIS): Accommodating error-components associated with a particular decision location in space, unobserved individual-specific heterogeneity in preferences (intrinsic biases) and in age/gender effects.	Simulator-based experiment with Austin area commuters in 2000.	Pseudo-random draws
	Srinivasan and Ramadurai (2006)	Error components and Random coefficients structure	Travel behavior and mode choice: Accommodating within-day dynamics and variations in mode-choice within and across individuals at the activity-episode level.	2000 RP multiday urban travel survey collected in the San Francisco Bay area.	Pseudo-random draws
MGEV	Bhat and Guo (2004)	Random coefficients with GEV base structure	Residential location choice: Allowing spatial correlation in adjacent spatial units due to unobserved location factors using a paired Generalized Nested Logit (GNL) structure, and unobserved individual-specific heterogeneity in responsiveness to travel time and other factors.	1996 RP urban travel survey from the Dallas-Fort Worth area.	Halton draws
	Bajwa *et al* (2006)	Nested logit with random coefficients structure	Joint departure time and mode choice: Accounting for correlation among alternative modes as well as the unobserved individual specific sensitivities to arrival time and other factors.	SP survey of commuters collected in Tokyo, Japan.	Information not provided
	Hess *et al.* (2004)	Nested and cross-nested logit with random coefficients structure	Mode choice: Accounting for inter-alternative correlation and random taste heterogeneity in travel time and alternative specific attributes.	1999 SP survey of mode choice collected in Switzerland.	Halton draws
	Lapparant (2006)	Nested logit with random coefficients structure	Mode choice: Accounting for correlation among alternative models as well as the unobserved individual-specific sensitivities to level-of-service and other factors	2001-2002 RP regional travel survey conducted in the French Parisian region of France.	Halton draws

	Srinivasan and Athuru (2005)	Nested logit with error components structure	Out-of-home maintenance participation: Accounting for correlation in solo participation, unobserved correlation between household members, and for correlation across episodes made by the same individual.	1996 RP urban travel survey collected in the San Francisco Bay Area.	Pseudo-random draws