CHAPTER 5: Flexible Model Structures for Discrete Choice Analysis

Chandra R. Bhat

The University of Texas at Austin

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 two classes of discrete choice models that relax one or more of the assumptions discussed above. The first class of models (here called "heteroscedastic models") are 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 (here called "flexible structure 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 third assumption 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 [for a detailed discussion of these procedures see Bhat (1997)].

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 parameters. The origin of these restrictions can be traced back to the requirement that the overall variance of alternatives be identical.

The rest of this chapter is in three sections. The heteroscedastic models and flexible structure models are discussed in Section 2 and 3. First, possible model formulations within the class are presented, and a preferred model formulation is then selected for further discussion. The structure of the preferred model structure is provided, followed by the estimation of the structure and a discussion of transport applications of the structure. Section 4 concludes the paper.

2 HETEROSCEDASTIC MODELS

2.1.   Model Formulations

The concept that heteroscedasticity in alternative error terms (*i.e.*, independent, but not identically distributed error terms) relaxes the IIA assumption is not new (Daganzo, 1979), but has received little (if any) attention in travel demand modeling and other fields. 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).

Daganzo (1979) used independent negative exponential distributions with different variances for the random-error components to develop a closed-form discrete choice model that does not have the IIA property. His 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 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. Then, random variation in the attributes that are defined only for the oddball alternative will generate increased variance in the overall random component of the oddball alternative relative to others. For example, operating schedule and fare structure define the utility of the transit alternative, but are not defined for other modal alternatives in a mode choice model. Consequently, measurement error in schedule and fare structure will contribute to the increased variance of transit relative to 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 has 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.

2.2.   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):

$$U_i = V_i + \varepsilon_i, \tag{1}$$

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 ith 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 ith alternative are:

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

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

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

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)

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

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

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

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

2.3.   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:

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

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

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

The expression within braces in the 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.* (1993)].

2.4.   Transport Applications

The HEV model has been applied to estimate discrete choice models based on revealed choice (RC) data as well as stated choice (SC) data.

The MNL, alternative nested logit structures, and the heteroscedastic model are estimated using RC data (Bhat, 1995) to examine the impact of improved rail service on intercity business travel in the Toronto-Montreal corridor. The nested logit structures are either inconsistent with utility maximization principles or are not significantly better than the MNL model. The HEV model, however, is found to be superior to the MNL model. The HEV model predicts smaller increases in rail shares and smaller decreases in non-rail shares than does the MNL in response to rail-service improvements. It also suggests a larger percentage decrease in air share and a smaller percentage decrease in auto share than does the MNL.

Hensher *et al.* (1999) applied the HEV model to estimate an intercity travel mode choice model from a combination of RC and SC choice data (they also discuss a latent-class HEV model in their paper that allows taste heterogeneity in a HEV model). The objective of the study was to identify the market for a proposed high-speed rail service in the Sydney-Canberra corridor. The revealed choice set includes four travel modes: air, car, bus or coach, and conventional rail. The stated choice set includes the four RC alternatives and the proposed high-speed rail alternative.

Hensher *et al.* estimate a pooled RC-SC model which accommodates scale differences between RC and SC data as well as scale differences among alternatives. The scale for each mode turns out to be about the same across the RC and SC datasets, possibly reflecting a well-designed stated choice task that captures variability levels comparable to actual revealed choices. Very interestingly, however, the scale for all non-car modes are about equal and substantially less than that of the car mode. This indicates much more uncertainty in the evaluation of non-car modes compared to the car mode.

Hensher (1997) has applied the HEV model in a related stated choice study to evaluate the choice of fare type for intercity travel in the Sydney-Canberra corridor conditional on the current mode used by each traveler. The current modes in the analysis include conventional train, charter coach, scheduled coach, air, and car. The projected patronage on a proposed high-speed rail mode is determined based on the current travel profile and alternative fare regimes.

Hensher (1998), in another effort, has applied the HEV model to the valuation of attributes (such as the value of travel time savings) from discrete choice models. Attribute valuation is generally based on the ratio of two or more attributes within utility expressions. However, using a common scale across alternatives can distort the relative valuation of attributes across alternatives. In Hensher's empirical analysis, the mean value of travel-time savings for public transport modes is much lower when an HEV model is used compared to a MNL model, because of confounding of scale effects with attribute parameter magnitudes. In a related and more recent study, Hensher (1999) applied the HEV model (along with other advanced models of discrete choice, such as the multinomial probit and mixed logit models, which we discuss later) to examine valuation of attributes for urban car-drivers.

Munizaga et al. (2000) evaluated the performance of several different model structures (including the HEV and the MNL model) in their ability to replicate heteroscedastic patterns across alternatives. They generated data with known heteroscedastic patterns for the analysis. Their results show that the MNL model does not perform well and does not provide accurate policy predictions in the presence of heteroscedasticity across alternatives, while the HEV model accurately recovers the target values of the underlying model parameters.

3. Flexible Structure 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 model structures 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.

3.1.   Model Formulations

Two closely related model formulations may be used to relax the IID (across alternatives) error structure and/or the assumption of response homogeneity: the multinomial probit (MNP) model and the mixed multinomial logit (MMNL) model.

The MNP structure allows a flexible structure for the covariance among the unobserved attributes of the alternatives. However, the MNP structure requires high-dimensional multivariate normal integration of the order of the number of alternatives in the choice probability expressions. Methodological developments in the past few years suggest approximating the high-dimensional integral with smooth, asymptotically unbiased, and efficient simulators. An MNP simulator that satisfies the above properties, and which provides strictly positive values for the choice probabilities, is a kernel-smoothed simulator with a MNL kernel. The use of the logit kernel simulator is equivalent to the addition of a Gumbel error term (distributed identically and independently across alternatives) to the utility of each alternative. Of course, the addition of such an error term changes the structure of the MNP model. But, as indicated formally by McFadden and Train (1997), the addition of such a Gumbel term is innocuous and does not change utility relationships. In fact, the model with the additional Gumbel error term can be made to approximate the model without the Gumbel term to any desired degree of closeness.

The estimation of the MNP model using a logit kernel-smoothed simulator takes the structure of a MMNL model since the unconditional choice probabilities are obtained by integrating the MNL form over the multivariate distribution of the random error terms in the MNP covariance matrix. This perspective of the MMNL model is useful since it highlights the point that there is no real substantive difference between the MNP and MMNL models (this issue is not always clear in the literature on discrete choice models).

Another issue in the estimation of the MNP model is the number of parameters to be estimated in the covariance matrix. A completely free covariance matrix (subject to certain

identification considerations) implies a general substitution pattern, but also generates a number of conceptual, statistical and practical problems, including difficulty in interpretation, highly non-intuitive model behavior, and low precision of covariance parameter estimates. Fortunately, in most instances, it is unnecessary to maintain a completely free covariance matrix because theoretical and intuitive considerations will imply restrictions on the covariance matrix. Even otherwise, it is possible to test rather general substitution patterns without specifying a completely free covariance matrix; this can be achieved by the use of carefully specified, and parsimonious, parametric forms for the error covariance matrix (see Ben-Akiva and Bolduc, 1996).

There are two ways that constraints implied by theoretical/intuitive considerations or parsimonious error covariance forms can be maintained in the MNP model. The first approach is to apply the constraints or the parsimonious forms directly on the MNP covariance matrix. The second approach is to use a factor-analytic approach with factors that generate the desired covariance structure, along with IID (across alternatives) error terms. In the first approach, the dimensionality of the integral in the choice probabilities is of the order of J-1, where J is the number of alternatives. In the second approach, the dimensionality is of the order of the number of factors (since it is conditional on the factors, the remainder error term is IID by specification). When the number of alternatives is large, as is the case in many choice contexts, the number of factors will be smaller than the number of alternatives. Thus, there can be substantial gains in simulator efficiency and estimation cost if one uses the factor-analytic approach (see also Brownstone and Train, 1999). Besides gains in computational efficiency, the factor-analytic approach explicitly requires the analyst to specify the structure underlying the substitution patterns among alternatives. From this perspective, it might be interpreted as a "structural" model, while imposing constraints on the MNP covariance matrix may be viewed as a "reduced-form" model.

To summarize, the use of a logit kernel (*i.e.*, the addition of an IID Gumbel error term in the utility of alternatives) leads to unbiased, smooth, and strictly positive choice probability simulators for the MNP model. The use of a factor-analytic formulation is equivalent to a structural specification and also leads to efficient simulators. A natural combination, then, is to use a logit kernel with a factor-analytic approach to realize the benefits of both specifications. In such a combined model formulation, the choice probabilities of the alternatives conditional on the factors takes the familiar MNL form. The unconditional choice probabilities are obtained by integrating the MNL form over the distribution of the random parameters in the factors. For this reason, the factor-

analytic approach with a logit kernel has a MMNL structure. A particularly desirable property of the factor-analytic MMNL structure (or simply the MMNL structure from hereon) is that it can approximate any discrete choice model derived from random utility maximization (including the MNP) as closely as one pleases. It also allows a non-normal distribution for the factors. Thus, the MMNL model is a very general discrete choice formulation. In the rest of this section, we focus on the MMNL structure.

3.2. MMNL Structure

The mixed MMNL model is a generalization of the well-known MNL model. It involves the integration of the MNL formula over the distribution of unobserved random parameters. It takes the structure

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta)\, f(\beta\,|\,\theta)\, d(\beta),\; L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qi}}}\; , \tag{11}$$

where 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 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).

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:

$$U_{qi} = \gamma' y_{qi} + \zeta_{qi}$$
$$\qquad = \gamma' y_{qi} + \mu' z_{qi} + \varepsilon_{qi}$$

(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 ($m = 1, 2, \ldots, M$) 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 are different. More general structures for $\mu' z_i$ in equation (12) are presented by Ben-Akiva and Bolduc (1996) and Brownstone and Train (1999).

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

$$U_{qi} = \beta'_q x_{qi} + \varepsilon_{qi}$$

(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).

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 (with cross-sectional data, the coefficients on the alternative-specific constants have to be considered as being deterministic).

Due to the equivalence between the random-coefficients and error-components formulations, and the more compact notation of the random-coefficients formulation, we will use the latter formulation is used in the discussion of the estimation methodology for the MMNL model in the next section.

3.3. MMNL Estimation Methodology

This section discusses the details of the estimation procedure for the random-coefficients MMNL model using each of three methods: the cubature method, the Pseudo-Monte Carlo (PMC) method, and the Quasi-Monte Carlo (QMC) method.

Considering equation (13) and separating out the effect of variables with fixed coefficients (including the alternative specific constant) from the effect of variables with random coefficients:

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

where α_{qi} is the effect of variables with fixed coefficients. Let $\beta_{qk} \sim N(\mu_k, \sigma_k)$, so that $\beta_{qk} = \mu_k + \sigma_k s_{qk}$ ($q = 1, 2, \ldots, Q$; $k = 1, 2, \ldots, K$). 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} ($q = 1, 2, \ldots, Q$; $k = 1, 2, \ldots, K$) is a standard normal variate. Further, let $V_{qi} = \alpha_{qi} + \sum_{k} \mu_k x_{qik}$.

The log (likelihood) function for the random-coefficients logit model may be written as

$$\langle = \sum_{q} \sum_{i} y_{qi} \log P_{qi}$$

$$= \sum_{q} \sum_{i} y_{qi} \log \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi}} + \sum_{k} \sigma_k s_{qk} x_{qik}}{\sum_{j} e^{V_{qj}} + \sum_{k} \sigma_k s_{qk} x_{qjk}} d\Phi(s_{q1}) d\Phi(s_{q2}) \ldots d\Phi(s_{qK}) \right\}, \tag{15}$$

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

$$y_{qi} = \begin{cases} 1 \text{ if the } q\text{th individual chooses alternative } i \\ (q = 1, 2, \ldots, Q, \ i = 1, 2, \ldots, I) \\ 0 \text{ otherwise}, \end{cases} \tag{16}$$

The cubature method, the PMC method, and the QMC method represent three different ways of evaluating the multidimensional integral involved in the log (likelihood) function.

Polynomial-based cubature method

To apply the cubature method, define $\varpi_k = s_{qk}/\sqrt{2}$ for all q. Then, the log (likelihood) function in equation (15) takes the following form:

$$
\langle \; = \sum_{q} \sum_{i} y_{qi} \log \left\{ \left(\frac{1}{\sqrt{\pi}} \right)^{K} \int_{\varpi_{1}=-\infty}^{\varpi_{1}=+\infty} \int_{\varpi_{2}=-\infty}^{\varpi_{2}=+\infty} \cdots \int_{\varpi_{K}=-\infty}^{\varpi_{K}=+\infty} \frac{e^{V_{qi}+\sqrt{2}\sum_{k}\sigma_{k}\varpi_{k}x_{qik}}}{\sum_{j} e^{V_{qj}+\sqrt{2}\sum_{k}\sigma_{k}\varpi_{k}x_{qjk}}} \times \right.
$$

$$
\left. e^{-\varpi_{1}^{2}} e^{-\varpi_{2}^{2}} \ldots e^{-\varpi_{K}^{2}} \, d\varpi_{1} d\varpi_{2} \ldots d\varpi_{K} \right\}.
$$

(17)

The above integration is now in an appropriate form for application of a multidimensional product formula of the Gauss-Hermite quadrature (see Stroud, 1971).

Random PMC method

This technique approximates the choice probabilities by computing the integrand in equation (15) at randomly chosen values for each s_{qk}. Since the s_{qk} terms are independent across individuals and variables, and are distributed standard normal, we generate a matrix *s* of standard normal random numbers with $Q*K$ elements (one element for each variable and individual combination) and compute the corresponding individual choice probabilities for a given value of the parameter vector to be estimated. This process is repeated *R* times for the given value of the parameter vector and the integrand is approximated by averaging over the computed choice probabilities in the different draws. This results in an unbiased estimator of the actual individual choice probabilities. The variance of $\langle$ decreases as *R* increases also has the appealing properties of being smooth (*i.e.*, twice differentiable) and being strictly positive for any realization of the finite *R* draws. The parameter vector is estimated as the vector value that maximizes the simulated log-likelihood function. Under rather weak regularity conditions, the PMC estimator is consistent, asymptotically efficient, and asymptotically normal. However, the estimator will generally be a biased simulation of the maximum likelihood (ML) estimator because of the logarithmic transformation of the choice probabilities in the log-likelihood function. The bias decreases with the variance of the probability simulator; that is, it decreases as the number of repetitions increase.

Random QMC method

The quasi-random Halton sequence is designed to span the domain of the S-dimensional unit cube uniformly and efficiently (the interval of each dimension of the unit cube is between 0 and 1). In one dimension, the Halton sequence is generated by choosing a prime number r $(r \geq 2)$ and expanding the sequence of integers 0, 1, 2, …g, ...G in terms of the base r:

$$g = \sum_{l=0}^{L} b_l r^l, \text{ where } 0 \leq b_1 \leq r-1 \text{ and } r^L \leq g \leq r^{L+1}. \tag{18}$$

Thus, g $(g = 1, 2, ..., G)$ can be represented by the *r-adic* integer string $b_l...b_1b_0$. The Halton sequence in the prime base r is obtained by taking the radical inverse of g $(g = 1, 2, ..., G)$ to the base r by reflecting through the radical point:

$$\varphi_r(g) = 0.b_0b_1...b_L(\text{base } r) = \sum_{l=0}^{L} b_l r^{-l-1}. \tag{19}$$

The sequence above is very uniformly distributed in the interval $(0,1)$ for each prime r. The Halton sequence in K dimensions is obtained by pairing K one-dimensional sequences based on K pairwise relatively prime integers (usually the first K primes):

$$\psi_g = [\varphi_{r1}(g), \varphi_{r2}(g),..., \varphi_{rs}(g)]. \tag{20}$$

The Halton sequence is generated number-theoretically rather than randomly and so successive points at any stage "know" how to fill in the gaps left by earlier points, leading to a uniform distribution within the domain of integration.

The simulation technique to evaluate the integral in the log-likelihood function of equation (15) involves generating the K-dimensional Halton sequence for a specified number of "draws" R for each individual. To avoid correlation in simulation errors across individuals, separate independent draws of R Halton numbers in K dimensions are taken for each individual. This is achieved by generating a Halton "matrix" **Y** of size G x K, where $G = R*Q+10$ (Q is the total number of individuals in the sample). The first ten terms in each dimension are then discarded because the integrand may be sensitive to the starting point of the Halton sequence. This leaves a $(R*Q)$ x K Halton matrix which is partitioned into Q sub-matrices of size R x K, each sub-matrix representing

the R Halton draws in K dimensions for each individual (thus, the first R rows of the Halton matrix $\mathbf{Y}$ are assigned to the first individual, the second R rows to the second individual, and so on).

The Halton sequence is uniformly distributed over the multi-dimensional cube. To obtain the corresponding multivariate normal points over the multi-dimensional domain of the real line, the inverse standard normal distribution transformation of $\mathbf{Y}$ is taken. By the integral transform result, $X = \Phi^{-1}(Y)$ provides the Halton points for the multi-variate normal distribution (see Fang and Wang, 1994; Chapter 4). The integrand in equation (15) is computed at the resulting points in the columns of the matrix $\mathbf{X}$ for each of the R draws for each individual and then the simulated likelihood function is developed in the usual manner as the average of the values of the integrand across the R draws.

Comparison of the estimation techniques

Bhat (1999a) recently proposed and introduced the use of the Halton sequence for estimating the mixed logit model and conducted Monte Carlo simulation experiments to study the performance of this QMC simulation method *vis-à-vis* the cubature and PMC simulation methods (this study, to the author's knowledge, is the first attempt at employing the QMC simulation method in discrete choice). Bhat's results indicate that the QMC method out-performs the polynomial-cubature and PMC methods for mixed logit model estimation. In ≤ 3 dimensions, simulation estimation with as few as 75 Halton draws provides considerably better accuracy than with 2000 pseudo-random draws. In higher dimensions (4 or 5), 100 Halton draws provide about the same level of accuracy as 2000 pseudo-random draws and 125 Halton draws provides much better accuracy in about one-tenth the time required for convergence using 2000 pseudo-random draws. Bhat also noted that this substantial reduction in computational cost has the potential to dramatically influence the use of the mixed logit model in practice. Specifically, given the flexibility of the mixed logit model to accommodate very general patterns of competition among alternatives and/or random coefficients, the use of the QMC simulation method of estimation should facilitate the application of behaviorally rich structures for discrete choice modeling. An application of the Halton sequence may be found in Bhat (1999b) in the context of a spatial model of work mode choice. A subsequent study by Train (1999) confirmed the substantial reduction in computational time for mixed logit estimation using the QMC method. More recently, Hensher (1999) has investigated Halton sequences involving

draws of 10, 25, 50, 100, 150 and 200, and compared the findings with random draws for mixed logit model estimation. He noted that the data fit and parameter values of the mixed logit model remain almost the same beyond 50 Halton draws. He concluded that the QMC method "is a phenomenal development in the estimation of complex choice models".

3.4.   Transport Applications

The transport applications of the MMNL model are discussed under two headings: error-components applications and random-coefficients applications.

Error-components applications

Brownstone and Train (1999) applied an error-components MMNL structure to model households' choices among gas, methanol, compressed natural gas (CNG), and electric vehicles. Their specification allows non-electric vehicles to share an unobserved random component, thereby increasing the sensitivity of non-electric vehicles to one another compared to an electric vehicle. Similarly, a non-CNG error component is introduced. Two additional error components related to the size of the vehicle are also introduced: one is a normal deviate multiplied by the size of the vehicle, and the second is a normal deviate multiplied by the luggage space. All these error components are statistically significant, indicating non-IIA competitive patterns.

Bhat (1998a) applied the MMNL model to a multidimensional choice situation. Specifically, his application accommodates unobserved correlation across both dimensions in a two-dimensional choice context. The model is applied to an analysis of travel-mode and departure-time choice for home-based social-recreational trips using data drawn from the 1990 San Francisco Bay area household survey. The empirical results underscore the need to capture unobserved attributes along both the mode and departure-time dimensions, both for improved data fit and for more realistic policy evaluations of transportation control measures.

Random-coefficients applications

There have been several applications of the MMNL model motivated from a random-coefficients perspective. Bhat (1998b) estimated a model of intercity travel-mode choice that accommodates variations in responsiveness to level-of-service measures due to both observed and unobserved individual characteristics. The model is applied to examine the impact of improved rail service on

weekday business travel in the Toronto-Montreal corridor. The empirical results show that not accounting adequately for variations in responsiveness across individuals leads to a statistically inferior data fit and also to inappropriate evaluations of policy actions aimed at improving inter-city transportation services.

Bhat (2000) formulated a MMNL model of multi-day urban travel-mode choice that accommodates variations in mode preferences and responsiveness to level-of-service. The model is applied to examine the travel-mode choice of workers in the San Francisco Bay area. Bhat's empirical results indicate significant unobserved variation (across individuals) in intrinsic mode preferences and level-of-service responsiveness. A comparison of the average response coefficients (across individuals in the sample) among the fixed-coefficient and random-coefficient models shows that the random-coefficients model implies substantially higher monetary values of time than the fixed-coefficient model. Overall, the empirical results emphasize the need to accommodate observed and unobserved heterogeneity across individuals in urban mode-choice modeling.

Train (1998) used a random-coefficients specification to examine the factors influencing anglers' choice of fishing sites. Explanatory variables in the model include fish stock (measured in fish per 1000 feet of river), esthetics rating of fishing site, size of each site, number of camp grounds and recreation access at site, number of restricted species at the site, and the travel cost to the site (including the money value of travel time). The empirical results indicate highly significant taste variation across anglers in the sensitivity to almost all the factors listed above. In this study, as well as the one by Bhat (2000), there was a very dramatic increase in data fit after including random variation in coefficients.

Mehndiratta (1997) proposed and formulated a theory to accommodate variations in the resource value of time in time-of-day choice for intercity travel. Mehndiratta then proceeded to implement his theoretical model using a random-coefficients specification for the resource value of disruption of leisure and sleep. He uses a stated choice sample in his analysis.

Hensher (2000) undertook a stated-choice analysis of the valuation of non-business travel time savings for car drivers undertaking long distance trips (up to 3 hours) between major urban areas in New Zealand. Hensher disaggregates overall travel time into several different components, including free flow travel time, slowed-down time, and stop time. The coefficients of each of these attributes are allowed to vary randomly across individuals in the population. The study showed significant taste heterogeneity to the various components of travel time, and adds to the

accumulating evidence that the restrictive travel time response homogeneity assumption undervalues the mean value of travel-time savings.

4. Conclusions

The structure, estimation techniques, and transport applications of two classes of discrete choice models — heteroscedastic models and flexible structure models—have been described, and within each class alternative formulations have been discussed. The formulations presented are quite flexible (this is especially the case with the flexible structure models), 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 model). 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 has now made the estimation of such analytically intractable model formulations very practical. In this regard, the recent use of QMC simulation techniques seems to be particularly effective, since it needs very few draws for accurate and efficient model estimation.

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 structures 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 structures 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 structure 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.

Acknowledgments

References

Ben-Akiva, M. and D. Bolduc (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. and S.R. Lerman (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. (1997) Recent methodological advances relevant to activity and travel behavior analysis, paper presented at: International Association of Travel Behavior Research Conference, Austin, Texas.

Bhat, C.R. (1998a) Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel mode and departure time choice, *Transportation Research B*, 32: 425-440.

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

Bhat, C.R. (1999a) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, Department of Civil Engineering, University of Texas.

Bhat, C.R. (1999b) A multi-level cross-classified model for discrete response variables, Department of Civil Engineering, University of Texas.

Bhat, C.R. (2000) Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling, *Transportation Science*, 34: 228-239.

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

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

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

Fang, K.-T and Y. Wang (1994) *Number-theoretic methods in statistics*. London: Chapman & Hall.

Hensher, D.A. (1997) A practical approach to identifying the market for high speed rail in the Sydney-Canberra corridor, *Transportation Research A*, 31: 431-446.

Hensher, D.A. (1998) Extending valuation to controlled value functions and non-uniform scaling with generalized unobserved variances, in T. Gärling, T. Laitila, and K. Westin, eds, *Theoretical foundations of travel choice modeling*. Oxford: Pergamon, 75-102.

Hensher, D.A. (1999) The valuation of travel time savings for urban car drivers: evaluating alternative model specifications, Institute of Transport Studies, University of Sydney, Technical Paper.

Hensher, D.A. (2000) Measurement of the valuation of travel time savings, *Journal of Transport Economics and Policy*, in press.

Hensher, D.A., J. Louviere and J. Swait (1999) Combining sources of preference data, *Journal of Econometrics*, 89: 197-221.

Luce, R. and P. Suppes (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 K. Train (1997) Mixed MNL models for discrete response, Department of Economics, University of California, Berkeley, CA, Working Paper.

Mehndiratta, S. (1997) Time-of-day effects in intercity business travel, PhD thesis, Department of Civil Engineering, University of California, Berkeley, CA.

Munizaga, M.A., B.G. Heydecker and J. Ortuzar (1999) Representation of heteroscedasticity in discrete choice models, *Transportation Research*, in press.

Press, W.H., S.A. Teukolsky and M. Nerlove (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 K. Train (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, 80, 647-657.

Stroud, A.H. (1971) *Approximate calculation of multiple integrals.* Englewood Cliffs, NJ: Prentice Hall.

Train, K. (1998) Recreation demand models with taste differences over people, *Land Economics*, 74: 230-239.

Train, K. (1999) Halton sequences for mixed logit, Department of Economics, University of California, Berkeley, CA, Technical Paper.