Random Utility-Based Discrete Choice Models for Travel Demand Analysis

Chandra R. Bhat

Department of Civil Engineering, ECJ Hall 6.810

The University of Texas at Austin, Austin, Texas 78712

Phone: 512-471-4535, Fax: 512-475-8744,

Email: bhat@mail.utexas.edu

1. INTRODUCTION

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. Before proceeding to review advanced discrete choice models, we first summarize the assumptions of the multinomial logit (MNL) formulation. 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 the utilities of all transit modes. As indicated by Koppelman and Sethi (2000), 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 whose values 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 (for example, 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 (that is, 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; see also Ben-Akiva and Lerman, 1985 for a detailed discussion of this property). Thus, relaxing the three assumptions may be important in many choice contexts.

In this chapter, we focus on three classes of discrete choice models which relax one or more of the assumptions discussed above and nest the multinomial logit model. The first class of models, which we will label as heteroscedastic models, 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, which we will refer to as Generalized Extreme Value (or GEV) models relax the independently distributed (across alternatives) assumptions, but do not relax the identically distributed assumption (part of the first assumption above) or

the assumptions of response homogeneity (second assumption). The third class of models, which we will label as 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. We do not focus on the third assumption implicit in the MNL model 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, 1997 for a detailed discussion of these procedures).

The rest of this paper is structured in three sections. Section 2 discusses heteroscedastic models. Section 3 focuses on GEV models. Section 4 presents flexible structure models. The final section concludes the paper. Within each of Sections 2, 3, and 4, the material is organized as follows. First, possible model formulations within that class are presented and a preferred model formulation is selected for further discussion. Next, the structure of the preferred model structure is provided, followed by the estimation of the structure, a brief discussion of transport applications of the structure, and a detailed presentation of results from a particular application of the structure in the travel behavior field.

2. HETEROSCEDASTIC MODELS

2.1 Model Formulations

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 which 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 multinomial logit 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 heteroscedastic extreme-value (HEV) model which assumes that the alternative error terms are distributed with a type I extreme value distribution. The variance 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 multinomial logit model and is flexible enough to allow differential cross-elasticities among all pairs of alternatives. In the rest of our discussion of heteroscedastic models, we will focus on the HEV model.

2.2 HEV Model Structure

The random utility of alternative i, U_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 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:

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

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 (P_i) from the set C of available alternatives:

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

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):

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

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 multinomial logit (the reader will note that the variance of the random error term ε_i of alternative i is equal to $U_i = V_i + \varepsilon_i$ where θ_i is the scale parameter).

The HEV model discussed above avoids the pitfalls of the IIA property of the multinomial logit 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 multinomial logit 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_{lk}) are characterized by the scale parameter of the random components of alternatives *i* and *j*:

$$\eta_{x_{lk}}^{P_i} > \eta_{x_{lk}}^{P_j} \text{ if } \theta_i < \theta_j; \quad \eta_{x_{lk}}^{P_i} = \eta_{x_{lk}}^{P_j} \text{ if } \theta_i = \theta_j; \quad \eta_{x_{lk}}^{P_i} < \eta_{x_{lk}}^{P_j} \text{ if } \theta_i > \theta_j \tag{6}$$

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 q^{th} individual and i^{th} alternative (we introduce the index for individuals 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:

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

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

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

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 $u = e^{-w}$. Then, $\lambda(w)dw = -e^{-u}du$ and $w = -\ln u$. Also define a function G_{qi} as:

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

Then we can re-write Equation (7) as

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

The expression within braces in the above equation 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; page 124).

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 multinomial logit, alternative nested logit structures, and the heteroscedastic model are estimated using RC data in Bhat (1995) to examine the impact of improved rail service on inter-city 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 multinomial logit model. The heteroscedastic extreme value model, however, is found to be superior to the multinomial logit model. The heteroscedastic model predicts smaller increases in rail shares and smaller decreases in non-rail shares than the multinomial logit 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 the multinomial logit.

Hensher *et al.* (1999) applied the HEV model to estimate an inter-city 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 this 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 data sets, 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 lesser 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 a 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 multinomial logit 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 multinomial logit 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.

2.5 Detailed Results From an Example Application

Bhat estimated the HEV model using data from a 1989 Rail Passenger Review conducted by VIA Rail (the Canadian national rail carrier). The purpose of the review was to develop travel demand models to forecast future intercity travel and estimate shifts in mode split in response to a variety of potential rail service improvements (including high-speed rail) in the Toronto-Montreal corridor (see KPMG Peat Marwick and Koppelman, 1990 for a detailed description of this data). Travel surveys were conducted in the corridor to collect data on intercity travel by four modes (car, air, train and bus). This data included socio-demographic and general trip-making characteristics of the traveler, and detailed information on the current trip (purpose, party size, origin and destination cities, *etc.*). The set of modes available to travelers

for their intercity travel was determined based on the geographic location of the trip. Level of service data were generated for each available mode and each trip based on the origin/destination information of the trip.

Bhat focused on intercity mode choice for paid business travel in the corridor. The study is confined to a mode choice examination among air, train, and car due to the very few number of individuals choosing the bus mode in the sample and also because of the poor quality of the bus data (see Forinash and Koppelman, 1993).

Five different models were estimated in the study: a multinomial logit model, three possible nested logit models, and the heteroscedastic extreme value model. The three nested logit models were: a) car and train (slow modes) grouped together in a nest which competes against air, b) train and air (common carriers) grouped together in a nest which competes against car, and c) air and car grouped together in a nest which competes against train. Of these three structures, the first two seem intuitively plausible, while the third does not.

The final estimation results are shown in Table 1 for the multinomial logit model, the nested logit model with car and train grouped as ground modes, and the heteroscedastic model. The estimation results for the other two nested logit models are not shown because the logsum parameter exceeded one in these specifications. This is not globally consistent with stochastic utility maximization (McFadden, 1978; Daly and Zachary, 1978).

A comparison of the nested logit model with the multinomial logit model using the likelihood ratio test indicates that the nested logit model fails to reject the multinomial logit model (equivalently, notice the statistically insignificance of the log sum parameter relative to a value of 1). However, a likelihood ratio test between the heteroscedastic extreme value model and the multinomial logit strongly rejects the multinomial logit in favor of the heteroscedastic specification (the test statistic is 16.56 which is significant at any reasonable level of significance when compared to a chi-squared statistic with two degrees of freedom). Table 1 also evaluates the models in terms of the adjusted likelihood ratio index ($\bar{\rho}^2$).[1] These values

[1] The adjusted likelihood ratio index is defined as follows:

$$\bar{\rho}^2 = 1 - \frac{L(M) - K}{L(C)}$$

again indicate that the heteroscedastic model offers the best fit in the current empirical analysis (note that the nested logit model and the heteroscedastic models can be directly compared to each other using the non-nested adjusted likelihood ratio index test proposed by Ben-Akiva and Lerman (1985); in the current case, the heteroscedastic model specification rejected the nested specification using this non-nested hypothesis test).

In the subsequent discussion on interpretation of model parameters, the focus will be on the multinomial logit and heteroscedastic extreme value models. The signs of all the parameters in the two models are consistent with *a priori* expectations (the car mode is used as the base for the alternative specific constants and alternative specific variables). The parameter estimates from the multinomial logit and the heteroscedastic model are also close to each other. However, there are some significant differences. The heteroscedastic model suggests a higher positive probability of choice of the train mode for trips which originate, end, or originate and end at a large city. It also indicates a lower sensitivity of travelers to frequency of service and travel cost; *i.e.*, the heteroscedastic model suggests that travelers place substantially more importance on travel time than on travel cost or frequency of service. Thus, according to the heteroscedastic model, reductions in travel time (even with a concomitant increase in fares) may be a very effective way of increasing the mode share of a travel alternative. The implied cost of in-vehicle travel time is $14.70 per hour in the multinomial logit and $20.80 per hour in the heteroscedastic model. The corresponding figures for out-of-vehicle travel time are $50.20 and $68.30 per hour, respectively.

The heteroscedastic model indicates that the scale parameter of the random error component associated with the train (air) utility is significantly greater (smaller) than that associated with the car utility (the scale parameter of the random component of car utility is normalized to one; the t-statistics for the train and scale parameters are computed with respect to a value of one). Therefore, the heteroscedastic model suggests unequal cross-elasticities among the modes.

Table 2 shows the elasticity matrix with respect to changes in rail level of service characteristics (computed for a representative inter-city business traveler in the corridor) for the multinomial logit and

---

where $L(M)$ is the model log-likelihood value, $L(C)$ is the log-likelihood value with only alternative specific constants and an IID error covariance matrix, and K is the number of parameters (besides the alternative specific constants) in the model.

heteroscedastic extreme value models.[2] Two important observations can be made from this table. First, the multinomial logit model predicts higher percentage decreases in air and car choice probabilities and a higher percentage increase in rail choice probability in response to an improvement in train level of service than the heteroscedastic model. Second, the multinomial logit elasticity matrix exhibits the IIA property because the elements in the second and third columns are identical in each row. The heteroscedastic model does not exhibit the IIA property; a one percent change in the level of service of the rail mode results in a larger percentage change in the probability of choosing air than auto. This is a reflection of the lower variance of the random component of the utility of air relative to the random component of the utility of car. We discuss the policy implications of these observations in the next section.

The observations made above have important policy implications at the aggregate level (these policy implications are specific to the Canadian context; caution must be exercised in generalizing the behavioral implications based on this single application). First, the results indicate that the increase in rail mode share in response to improvements in the rail mode is likely to be substantially lower than what might be expected based on the multinomial logit formulation. Thus, the multinomial logit model overestimates the potential ridership on a new (or improved) rail service and, therefore, overestimates revenue projections. Second, the results indicate that the potential of an improved rail service to alleviate auto-traffic congestion on intercity highways and air-traffic congestion at airports is likely to be lesser than that suggested by the multinomial logit model. This finding has a direct bearing on the evaluation of alternative strategies to alleviate intercity travel congestion. Third, the differential cross-elasticities of air and auto modes in the heteroscedastic logit model suggests that an improvement in the current rail service will alleviate air-traffic congestion at airports more so than alleviating auto-congestion on roadways. Thus, the potential benefit from improving the rail service will depend on the situational context; that is, whether the thrust of the congestion-alleviation effort is to reduce roadway congestion or to reduce air traffic congestion. These findings point to the deficiency of the multinomial logit model as a tool to making informed policy decisions to alleviate intercity travel congestion in the specific context of Bhat's application.

---

[2] Since the objective of the original study for which the data were collected was to examine the effect of alternative improvements in rail level of service characteristics, we focus on the elasticity matrix corresponding to changes in rail level of service here.

3. THE GEV CLASS OF MODELS

The GEV class of models relaxes the IID assumption of the MNL by allowing the random components of alternatives to be correlated, while maintaining the assumption that they are identically distributed (*i.e.*, identical, non-independent random components). This class of models assumes a type I extreme value (or Gumbel) distribution for the error terms. All the models belonging to this class nest the multinomial logit and result in closed-form expressions for the choice probabilities. In fact, the MNL is also a member of the GEV class, though we will reserve the use of the term "GEV class" to those models that constitute generalizations of the MNL.

The general structure of the GEV class of models was derived by McFadden (1978) from the random utility maximization hypothesis, and generalized by Ben-Akiva and Francois (1983). Several specific GEV structures have been formulated and applied within the GEV class, including the Nested Logit (NL) model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978), the Paired Combinatorial Logit (PCL) model (Chu, 1989; Koppelman and Wen, 2000), the Cross-Nested Logit (CNL) model (Vovsha, 1997), the Ordered GEV (OGEV) model (Small, 1987), the Multinomial Logit-Ordered GEV (MNL-OGEV) model (Bhat, 1998c), and the Product Differentiation Logit (PDL) model (Bresnahan *et al*., 1997), and the Generalized Nested Logit (GNL) model (Wen and Koppelman, 2001).

The nested logit (NL) model permits covariance in random components among subsets (or nests) of alternatives (each alternative can be assigned to one and only one nest). Alternatives in a nest exhibit an identical degree of increased sensitivity relative to alternatives not in the nest (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978). Each nest in the NL structure has associated with it a dissimilarity (or logsum) parameter that determines the correlation in unobserved components among alternatives in that nest (see Daganzo and Kusnic, 1993). The range of this dissimilarity parameter should be between 0 and 1 for all nests if the NL model is to remain globally consistent with the random utility maximizing principle. A problem with the NL model is that it requires *a priori* specification of the nesting structure. This requirement has at least two drawbacks. First, the number of different structures to estimate in a search for the best structure increases rapidly as the number of alternatives increases. Second, the actual competition structure among alternatives may be a continuum that cannot be accurately represented by partitioning the alternatives into mutually exclusive subsets.

The paired combinatorial logit (PCL) model initially proposed by Chu (1989) and recently examined in detail by Koppelman and Wen (2000) generalizes, in concept, the nested logit model by allowing differential correlation between each pair of alternatives (the nested logit model, however, is not nested within the PCL structure). Each pair of alternatives in the PCL model has associated with it a dissimilarity parameter (subject to certain identification considerations that Koppelman and Wen are currently studying) that is inversely related to the correlation between the pair of alternatives. All dissimilarity parameters have to lie in the range of 0 to 1 for global consistency with random utility maximization.

Another generalization of the nested logit model is the cross-nested logit (CNL) model of Vovsha (1997). In this model, an alternative need not be exclusively assigned to one nest as in the nested logit structure. Instead, an alternative can appear in different nests with different probabilities based on what Vovsha refers to as allocation parameters. A single dissimilarity parameter is estimated across all nests in the CNL structure. Unlike in the PCL model, the nested logit model can be obtained as a special case of the CNL model when each alternative is unambiguously allocated to one particular nest. Vovsha proposes a heuristic procedure for estimation of the CNL model. This procedure appears to be rather cumbersome and its heuristic nature makes it difficult to establish the statistical properties of the resulting estimates.

The ordered GEV model was developed by Small (1987) to accommodate correlation among the unobserved random utility components of alternatives close together along a natural ordering implied by the choice variable (examples of such ordered choice variables might include car ownership, departure time of trips, *etc.*). The simplest version of the OGEV model (which Small refers to as the standard OGEV model) accommodates correlation in unobserved components between the utilities of each pair of adjacent alternatives on the natural ordering; that is, each alternative is correlated with the alternatives on either side of it along the natural ordering.[3] The standard OGEV model has a dissimilarity parameter that is inversely related to the correlation between adjacent alternatives (this relationship does not have a closed form, but the correlation implied by the dissimilarity parameter can be obtained numerically). The dissimilarity parameter has to lie in the range of 0 to 1 for consistency with random utility maximization.

---

[3] The reader will note that the nested logit model cannot accommodate such a correlation structure because it requires alternatives to be grouped into mutually exclusive nests.

The MNL-OGEV model formulated by Bhat (1998c) generalizes the nested logit model by allowing adjacent alternatives within a nest to be correlated in their unobserved components. This structure is best illustrated with an example. Consider the case of a multi-dimensional model of travel mode and departure time for nonwork trips. Let the departure time choice alternatives be represented by several temporally contiguous discrete time periods in a day such as a.m. peak (6a.m.-9a.m.), a.m. mid-day (9a.m.-12 noon), p.m. mid-day (12 noon-3p.m.), p.m. peak (3p.m.-6p.m.), and other (6p.m.-6a.m.). An appropriate nested logit structure for the joint mode-departure time choice model may allow the joint choice alternatives to share unobserved attributes in the mode choice dimension, resulting in an increased sensitivity among time-of-day alternatives of the same mode relative to the time-of-day alternatives across modes. However, in addition to the uniform correlation in departure time alternatives sharing the same mode, there is likely to be increased correlation in the unobserved random utility components of each pair of adjacent departure time alternatives due to the natural ordering among the departure time alternatives along the time dimension. Accommodating such a correlation generates an increased degree of sensitivity between adjacent departure time alternatives (over and above the sensitivity among non-adjacent alternatives) sharing the same mode. A structure that accommodates the correlation patterns just discussed can be formulated by using the multinomial logit (MNL) formulation for the higher-level mode choice decision and the standard ordered generalized extreme-value (OGEV) formulation (see Small, 1987) for the lower-level departure time choice decision (*i.e.*, the MNL-OGEV model).

More recently, Wen and Koppelman (2001) proposed a general GEV model structure, which they referred to as the General Nested Logit (GNL) model. Swait (2001), independently, proposed a similar structure, which he refers to as the choice set Generation Logit (GenL) model; Swait's derivation of the GenL model is motivated from the concept of latent choice sets of individuals, while Wen and Koppelman's derivation of the GNL model is motivated from the perspective of flexible substitution patterns across alternatives. Wen and Koppelman (2001) illustrate the general nature of the GNL model formulation by deriving the other GEV model structures mentioned earlier as special restrictive cases of the GNL model or as approximations to restricted versions of the GNL model.

The GNL model is conceptually appealing because it is a very general structure and allows substantial flexibility. However, in practice, the flexibility of the GNL model can be realized only if one is

able and willing to estimate a large number of dissimilarity and allocation parameters. The net result is that the analyst will have to impose informed restrictions on the general GNL model formulation that are customized to the application context under investigation.

The advantage of all the GEV models discussed above is that they allow relaxations of the independence assumption among alternative error terms while maintaining closed-form expressions for the choice probabilities. The problem with 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 variance of the joint alternatives be identical in the GEV models. In addition, the GEV models do not relax the response homogeneity assumption discussed in the previous section.

In the rest of the discussion on GEV models, we will focus on the GNL model since it subsumes other GEV models proposed to date as special cases.

3.2 GNL Model Structure

The GNL model can be derived from the GEV postulate using the following function:

$$G = (y_1, y_2, \ldots, y_I) = \sum_m \left[\sum_{i' \in N_m} (\alpha_{i'm}\, y_{i'})^{1/\rho_m} \right]^{\rho_m}, \tag{11}$$

where N_m is the set of alternatives belonging to nest m, represents an allocation parameter characterizing the portion of alternative i assigned to nest m ($0 < \alpha_{im} < 1$; $\sum_m \alpha_{im} = 1 \; \forall \; i$), and ρ_m is a dissimilarity parameter for nest m ($0 < \rho_m \; ? \; 1$). Then it is easy to verify that G is non-negative, homogenous of degree one, tending toward $+\infty$ when any argument y_i tends toward $+\infty$, and whose n^{th} non partial derivatives are non-negative for odd n and non-positive for even n because $0 < \rho_m < 1$. Thus the following function represents a cumulative extreme-value distribution:

$$F = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_I) = \exp\left\{ -\sum_m \left[\sum_{i' \in N_m} (\alpha_{i'm}\, \rho^{-\varepsilon_{i'}})^{1/\rho_m} \right]^{\rho_m} \right\}, \tag{12}$$

To obtain the probability of choice for each alternative i in the GNL model, consider a utility maximizing decision process where the utility of each alternative i (U_i) is written in the usual form as the sum of a deterministic component (V_i) and a random component E_i . If the random components follow the CDF in Equation (12), then, by the GEV postulate, the probability of choosing the i^{th} alternative is:

$$P_i = \frac{\sum_m \left[(\alpha_{im} e^{V_i})^{1/\rho_m} \left(\sum_{i' \in N_m} (\alpha_{i'm} e^{V_{i'}})^{1/\rho_m} \right)^{\rho_m - 1} \right]}{\sum_m \left(\sum_{i' \in N_m} (\alpha_{i'm} e^{V_{i'}})^{1/\rho_m} \right)^{\rho_m}} \tag{13}$$

The cross elasticity of a pair of alternatives i and j, which appear in one or more common nests, is

$$\eta_{x_i}^{P_j} = -\left[P_i + \frac{\sum_m \left(\frac{1}{\rho_m} - 1 \right) P_m P_{i|m} P_{j|m}}{P_j} \right] \beta' x_i \tag{14}$$

If the two alternatives i and j do not appear in any common nest, the cross-elasticity reduces to zero. Wen and Koppelman also demonstrate that the correlation between two alternatives i and j is a function of both the allocation parameters and the dissimilarity parameters.

3.3 GNL Model Estimation

The GNL model may be estimated using the commonly-used maximum likelihood method. The parameter to be estimated in the GNL structure include variable coefficients, the dissimilarity parameters ρ_m (m=1,2,...,M), and the allocation parameters α_{im} , i=1,2,...I, m=1,2,...M). All the dissimilarity and allocation parameters need to be between 0 and 1, and the allocation parameters for each alternative should sum to 1. Wen and Koppelman used a constrained maximum likelihood procedure to estimate the model. It should be noted that the maximum number of dissimilarity parameters that can be estimated is one less than the number of pairs of alternatives.

3.4 GNL Model Applications

The GNL model was proposed recently by Wen and Koppelman. The results of their application are discussed in detail in the next section. In most practical situations, the analyst will have to impose informed restrictions on the GNL formulation. Such restrictions might lead to models such as the PCL, the OGEV, the MNL-OGEV, and the CNL models. In addition, the NL model can also be shown to be essentially the same as a restricted version of the GNL. Since there have been several applications of the NL model, and we have reviewed studies that have used the other GEV structures, we proceed to a detailed presentation of the GNL model by Wen and Koppelman.

3.5 Detailed Results From an Application of the GNL Model

Wen and Koppelman use the same Canadian rail data set used by Bhat (1995) and discussed in Section 2.5. They examined intercity mode choice in the Toronto-Montreal corridor. The universal choice set includes air, train, bus, and car.

Table 3 shows the results that Wen and Koppelman obtained from the GNL model and the MNL model. Wen and Koppelman also estimated several NL structures, a PCL model, and CNL models. However, the GNL model provided better data fit in their application context.

Table 3 provides the expected impacts of the level of service variables. The table also indicates that the model parameters tend to be smaller in magnitude in the GNL model relative to the MNL. However, the values of time are about the same from the two models. Most importantly, the differences in coefficient between the two models, combined with the correlation patterns generated by the GNL model, are likely to produce different mode share forecasts in response to policy actions or investment decisions.

4. FLEXIBLE STRUCTURE MODELS

The HEV and GEV class of models 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 which 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. This section focuses on model structures that explicitly nest the MNL model.

4.1 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 mixed multinomial logit (MMNL) model and the mixed GEV (MGEV) model.

The mixed multinomial logit (MMNL) model is a generalization of the well-known multinomial logit (MNL) model. It involves the integration of the multinomial logit formula over the distribution of unobserved random parameters. It takes the structure shown below:

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta \mid \theta) d(\beta), \; L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_{j} e^{\beta' x_{qj}}}, \tag{15}$$

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 (15) 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), as we discuss later in Section 4.2 (of course, the MMNL structure can also accommodate both a non IID error structure across alternatives as well as response heterogeneity).

The MGEV class of models use a GEV model as a core, and superimposes a mixing distribution on the GEV core to accommodate response heterogeneity and/or additional heteroscedasticity/correlation across alternative error terms. A question that arises here is as follows: Why would one want to consider an MGEV model structure when a MMNL model can already capture response heterogeneity and any identifiable pattern of heteroscedasticity/ correlation across alternative error

terms? That is, why would one want to consider a GEV core to generate a certain inter-alternative error correlation pattern when such a correlation pattern can be generated as part of a MMNL model structure? Bhat and Guo (2002) provide situations where an MGEV model may be preferred to an equivalent MMNL model. Consider, for instance, a 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 alternatives that are contiguous to be correlated. In the MMNL structure, such a correlation structure will require the specification of as many error components as the number of pairs of spatially-contiguous alternatives. In a residential choice context, the number of error components to be specified will therefore be very large (in the 100s or 1000s). This will require the computation of very high dimensional integrals (in the order of 100s of 1000s) in the MMNL structure. On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation. However, the GEV model structure cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such heterogeneity, leading to a parsimonious and powerful MGEV structure.

In the rest of this section, we will focus on the MMNL model structure, since all the concepts and techniques for the MMNL model are readily transferable to the MGEV model structure.

4.2 MMNL Model Structure

In this section, we discuss the MMNL structure from an error components viewpoint as well as from a random-coefficient viewpoint.

4.2.1 Error-components structure

The error components structure partitions the overall random term associated with each alternative's utility into two components: one component which allows the unobserved error terms to be non-identical and non-independent across alternatives, and the other which 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:

$$U_{qi} = \gamma' y_{qi} + \zeta_{qi} = \gamma' y_{qi} + \mu' z_{qi} + \epsilon_{qi} \quad (16)$$

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 whose elements 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 to allow 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 are different. More general structures for $\mu' z_i$ in Equation (16) are presented by Ben-Akiva and Bolduc (1996) and Brownstone and Train (1999).

4.2.2 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:

$$U_{qi} = \beta_q' x_{qi} + \epsilon_{qi} \quad (17)$$

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 (15). 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 beta 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).

The reader will note that the error-components specification in Equation (16) and the random-coefficients specification in Equation (17) are structurally equivalent. Specifically, if β_q is distributed with a mean of γ and deviation μ, then Equation (17) is identical to Equation (16) with $x_{qi} = y_{qi} = z_{qi}$. However, this apparent restriction for equality of Equations (16) and (17) is purely notational. Elements of x_{qi} that do not appear in z_{qi} can be viewed as variables whose coefficients are deterministic in the population, while elements of x_{qi} that do not enter in y_{qi} may be viewed as variables whose coefficients 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 in the discussion of the estimation methodology for the mixed logit model in the next section.

4.3 MMNL Estimation Methodology

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

Consider Equation (17) and separate out the effect of variables with fixed coefficients (including the alternative specific constant) from the effect of variables with random coefficients:

$$U_{qi} = \alpha_{qi} + \sum_{k=1}^{K} \beta_{qk} x_{qik} + \epsilon_{qi}, \tag{18}$$

where α_{qi} is the effect of variables with fixed coefficients. Let $\beta_{qk} \sim N(\mu_k, \sigma_k)$, so that $\beta_{qk} = \mu_k + \sigma_k \varepsilon_{qk}$ $(q = 1,2,\ldots,Q;\ k = 1,2,\ldots,K)$. In this notation, we are implicitly assuming that the β_{qk}'s 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). $\varepsilon_{qk}\,(q=1,2,\ldots,Q;\ k=1,2,\ldots,K)$ 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:

$$LL=\sum_q\sum_i y_{qi}\log P_{qi}=\sum_q\sum_i y_{qi}\log\left\{\int_{\varepsilon_{q1}=-\infty}^{\varepsilon_{q1}=+\infty}\int_{\varepsilon_{q2}=-\infty}^{\varepsilon_{q2}=+\infty}\cdots\int_{\varepsilon_{qK}=-\infty}^{\varepsilon_{qK}=+\infty}\frac{e^{V_{qi}+\sum_k \sigma_k \varepsilon_{qk} x_{qik}}}{\sum_j e^{V_{qj}+\sum_k \sigma_k \varepsilon_{qk} x_{qjk}}}\,d\Phi(\varepsilon_{q1})\,d\Phi(\varepsilon_{q2})\cdots d\Phi(\varepsilon_{qK})\right\}, \tag{19}$$

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

$$y_{qi}=\begin{cases}1 \text{ if the } q\text{th individual chooses alternative } i\\ 0 \text{ otherwise,}\end{cases}\qquad (q=1,2,\ldots,Q,\ i=1,2,\ldots,I) \tag{20}$$

The cubature method, the Pseudo-Monte Carlo (PMC) method, and the Quasi-Monte Carlo (QMC) method represent three different ways of evaluating the multi-dimensional integral involved in the log-likelihood function.

4.3.1 Polynomial-based cubature method

To apply the cubature method, define $\omega_k=\varepsilon_{qk}/\sqrt{2}$ for all q. Then, the log-likelihood function in Equation (19) takes the following form:

$$LL=\sum_q\sum_i y_{qi}\log\left\{\left(\frac{1}{\sqrt{\pi}}\right)^K\int_{\omega_1=-\infty}^{\omega_1=+\infty}\int_{\omega_2=-\infty}^{\omega_2=+\infty}\cdots\int_{\omega_K=-\infty}^{\omega_K=+\infty}\frac{e^{V_{qi}+\sqrt{2}\sum_k \sigma_k \omega_k x_{qik}}}{\sum_j e^{V_{qj}+\sqrt{2}\sum_k \sigma_k \omega_k x_{qjk}}}e^{-\omega_1^2}e^{-\omega_2^2}\cdots e^{-\omega_K^2}\,d\omega_1\,d\omega_2\cdots d\omega_K\right\}. \tag{21}$$

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

4.3.2 Pseudo-random Monte Carlo (PMC) method

This technique approximates the choice probabilities by computing the integrand in Equation (19) at randomly chosen values for each ε_{qk} . Since the ε_{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. It's variance decreases as *R* increases. It 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.

4.3.3 Quasi-random Monte Carlo (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*>=2) and expanding the sequence of integers 0,1,2,...,*g*,...,*G* in terms of the base *r*:

$$g = \sum_{l=0}^{L} b_l r^l, \text{ where } 0 \le b_l \le r-1 \text{ and } r^L \le g < r^{L+1}. \qquad (22)$$

Thus, *g* (*g*=1,2,...*G*) can be represented by the r-*adic* integer string $b_l \ldots b_1 b_0$. 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:

$$\varphi_r(g) = 0.b_0 b_1 \ldots b_L (\text{base } r) = \sum_{l=0}^{L} b_l r^{-l-1} \tag{23}$$

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):

$$\psi_g = (\varphi_{r_1}(g), \varphi_{r_2}(g), \ldots, \varphi_{r_s}(g)) \tag{24}$$

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 (19) 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 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 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 (19) is computed at the resulting points in the columns of the matrix 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.

Bhat (2001) 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 quasi-Monte Carlo (QMC) simulation method vis-a-vis the cubature and pseudo-Monte Carlo (PMC) simulation methods (this study, to the author's knowledge, is the first attempt at employing the QMC simulation method in discrete choice literature). Bhat's results indicate that the QMC method out-performs the polynomial-cubature and pseudo-Monte Carlo (PMC) methods for mixed logit model estimation. Bhat notes 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. Another subsequent study by Train (1999) confirms the substantial reduction in computational time for mixed logit estimation using the QMC method. Hensher (1999) has also investigated Halton sequences and compared the findings with random draws for mixed logit model estimation. He notes that the data fit and parameter values of the mixed logit model remain almost the same beyond 50 Halton draws. He concludes that the quasi-Monte Carlo method "is a phenomenal development in the estimation of complex choice models".

4.3.4 Scrambled and randomized QMC method

Bhat (2002) notes that a problem with the Halton sequence is that there is strong correlation between higher coordinates of the sequence. This is because of the cycles of length r for the prime r. Thus, when two large prime-based sequences, associated with two high dimensions, are paired, the corresponding unit square face of the S-dimensional cube is sampled by points that lie on parallel lines. For example, the fourteenth dimension (corresponding to the prime number 43) and the fifteenth dimension (corresponding to the prime number 47) consist of 43 and 47 increasing numbers, respectively. This generates a correlation between the fourteenth and fifteenth coordinates of the sequence. This is illustrated diagrammatically in the first plot of Figure 1. The consequence is a rapid deterioration in the uniformity of the Halton sequence in high dimensions (the deterioration becomes clearly noticeable beyond five dimensions).

Number theorists have proposed an approach to improve the uniformity of the Halton sequence in high dimensions. The basic method is to break the correlations between the coordinates of the standard Halton sequence by scrambling the cycles of length r for the prime r. This is accomplished by permutations of the coefficients b_l in the radical inverse function of Equation (23). The resulting scrambled Halton sequence for the prime r is written as:

$$\varphi_r(g) = \sum_{l=0}^{L} \sigma_r(b_l(g)) r^{-l-1}, \tag{25}$$

where σ_r is the operator of permutations on the digits of the expansion $b_l(g)$ (the standard Halton sequence is the special case of the scrambled Halton sequence with no scrambling of the digits $b_l(g)$). Different researchers (see Braaten and Weller, 1979; Hellekalek, 1984; Kocis and Whiten, 1997) have suggested different algorithms for arriving at the permutations of the coefficients b_l in Equation (25). The permutations used by Braaten and Weller are presented in the Appendix for the first ten prime numbers. Braaten and Weller have also proved that their scrambled sequence retains the theoretically appealing N^{-1} order of integration error of the standard Halton sequence.

An example would be helpful in illustrating the scrambling procedure of Braaten and Weller. These researchers suggest the following permutation of (0,1,2) for the prime 3: (0,2,1). As indicated earlier, the 5^{th} number in base 3 of the Halton sequence in digitized form is 0.21. When the permutation above is applied, the 5^{th} number in the corresponding scrambled Halton sequence in digitized form is 0.21, which when expanded in base 3 translates to $1 \times 3^{-1} + 2 \times 3^{-2} = 5/9$. The first 8 numbers in the scrambled sequence corresponding to base 3 are 2/3, 1/3, 2/9, 8/9, 5/9, 1/9, 7/9, 4/9.

The Braaten and Weller method involves different permutations for different prime numbers. As a result of this scrambling, the resulting sequence does not display strong correlation across dimensions as does the standard Halton sequence. This is illustrated in the second plot of Figure 1, which plots 150 scrambled Halton points in the fourteenth and fifteenth dimensions. A comparison of the two plots in Figure 1 clearly indicates the more uniform coverage of the scrambled Halton sequence relative to the standard Halton sequence.

In addition to the scrambling of the standard Halton sequence, Bhat also suggests a randomization procedure for the Halton sequence based on a procedure developed by Tuffin (1996). The randomization is useful because all QMC sequences (including the standard Halton and scrambled Halton sequences discussed above) are fundamentally deterministic. This deterministic nature of the sequences does not permit the practical estimation of the integration error. Theoretical results exist for estimating the integration error, but these are difficult to compute and can be very conservative.

The essential concept of randomizing QMC sequences is to introduce randomness into a deterministic QMC sequence that preserves the uniformly distributed and equidistribution properties of the underlying QMC sequence (see Shaw, 1988; Tuffin, 1996). One simple way to introduce randomness is based on the following idea. Let $\psi^{(N)}$ be a QMC sequence of length N over the S-dimensional cube $\{0,1\}^S$ and consider any S-dimensional uniformly distributed vector in the S-dimensional cube (u ?$\{0,1\}^S$). $\psi^{(N)}$ is a matrix of dimension N x S, and u is a vector of dimension 1xS. Construct a new sequence $\chi^{(N)} = \{\psi^{(N)} + u \otimes 1^{(N)}\}$, where {.} denotes the fractional part of the matrix within parenthesis, $\otimes$ represents the kronecker or tensor product, and $1^{(N)}$ is a unit column vector of size N (the kronecker product multiplies each element of u with the vector $1^{(N)}$). The net result is a sequence $\chi^{(N)}$ whose elements χ_{ns} are obtained as $\psi_{ns} + u_s$ if $\psi_{ns} + u_s \le 1$, and $\psi_{ns} + u_s - 1$ if $\psi_{ns} + u_s > 1$. It can be shown that the sequence $\chi^{(N)}$ so formed is also a QMC sequence of length N over the S-dimensional cube $\{0,1\}^S$. Tuffin provides a formal proof for this result, which is rather straightforward but tedious. Intuitively, the vector u simply shifts the points of each coordinate of the original QMC sequence $\psi^{(N)}$ by a certain value. Since all the points within each coordinate are shifted by the same amount, the new sequence will preserve the equidistribution property of the original sequence. This is illustrated in Figure 2 in two dimensions. The first diagram in Figure 2 plots 100 points of the standard Halton sequence in the first two dimensions. The second diagram plots 100 points of the standard Halton sequence shifted by 0.5 in the first dimension and 0 in the second dimension. The result of the shifting is as follows. For any point below 0.5 in the first dimension in the first diagram (for example, the point marked 1), the point gets moved by 0.5 toward the right in the second diagram. For any point above 0.5 in the first dimension in the first diagram (such as the point marked 2), the point gets moved to the right, hits the right edge, bounces off this edge to the left edge, and is carried forward so that the total

distance of the shift is 0.5 (another way to visualize this shift is to transform the unit square into a cylinder with the left and right edges "sewn" together; then the shifting entails moving points along the surface of the cylinder and perpendicular to the cylinder axis). Clearly, the two-dimensional plot in the second diagram of Figure 2 is also well-distributed because the relative positions of the points do not change from that in Figure 1; there is simply a shift of the overall pattern of points. The last diagram in Figure 2 plots the case where there is a shift in both dimensions; 0.5 in the first and 0.25 in the second. For the same reasons discussed in the context of the shift in one dimension, the sequence obtained by shifting in both dimensions is also well-distributed.

It should be clear from above that any vector $u ? \{0,1\}^S$ can be used to generate a new QMC sequence from an underlying QMC sequence. An obvious way of introducing randomness is then to randomly draw u from a multidimensional uniform distribution.

An important point to note here is that randomizing the standard Halton sequence as discussed earlier does not break the correlations in high dimensions because the randomization simply shifts all points in the same dimension by the same amount. Thus, randomized versions of the standard Halton sequence will suffer from the same problems of non-uniform coverage in high dimensions as the standard Halton sequence. To resolve the problem of non-uniform coverage in high dimensions, the scrambled Halton sequence needs to be used.

Once a scrambled and randomized QMC sequence is generated, Bhat proposes a simulation approach for estimation of the mixed logit model that is similar to the standard Halton procedure discussed on the previous section.

4.3.5 Bayesian estimation of MNL

Some recent papers (Brownstone, 2000; Train, 2001) have considered a Bayesian estimation approach for MMNL model estimation as opposed to the classical estimation approaches discussed above. The general results from these studies appear to suggest that the classical approach is faster when mixing distributions with bounded support such as triangulars are considered, or when there is a mix of fixed and random coefficients in the model. On the other hand, the Bayesian estimation appears to be faster when considering the normal distribution and its transformations, and when all coefficients are random and are

correlated with one another. However, in the overall, the results suggest that the choice between the two estimation approaches should depend more on interpretational ease in the empirical context under study rather than computational efficiency considerations.

4.4 Transport Applications

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

4.4.1 Error-components applications

Brownstone and Train (1999) applied an error-components mixed multinomial logit structure to model households' choices among gas, methanol, compressed natural gas (CNG), and electric vehicles, using stated choice (SC) data collected in 1993 from a sample of households in California. Brownstone and Train allow 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.

Brownstone *et al.* (2000) extended the analysis of Brownstone and Train (1999) to estimate a model of choice among alternative-fuel vehicles using both stated choice and revealed choice (RC) data. The RC data was collected about 15 months after the SC data, and recorded actual vehicle purchase behavior since the collection of the SC data. Brownstone *et al.* (2000) maintain the error-components structure developed in their earlier study, and also accommodate scale differences between RC and SC choices.

Bhat (1998a) applied the mixed multinomial logit (MMNL) model to a multi-dimensional 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 as well as for more realistic policy evaluations of transportation control measures.

4.4.2 Random-coefficients applications

There have been several applications of the mixed multinomial logit model motivated from a random-coefficients perspective.

Bhat (1998b) estimated a model of inter-city 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 mixed multinomial logit 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), aesthetics 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 Bhat's (2000) study, there is 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 inter-city 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) undertakes a stated choice analysis of the valuation of non-business travel time savings for car drivers undertaking long distance trips (up to three 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 finds 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.

In addition to the studies identified in Sections 4.4.1 and 4.4.2, some recent studies have included both inter-alternative error correlations (in the spirit of an error-components structure) as well as unobserved heterogeneity among decision-making agents (in the spirit of the random coefficients structure). Such studies include Hensher and Greene (2000), Bhat and Castelar (2002), and Han and Algers (2001).

4.5 Detailed Results From an Example Application

Bhat uses an error components motivation for the analysis of mode and departure time choice for social-recreational trips in the San Francisco Bay area. Bhat suggests the use of a MMNL model to accommodate unobserved correlation in error terms across both the model and temporal dimension simultaneously. The data for this study are drawn from the San Francisco Bay Area Household Travel Survey conducted by the Metropolitan Transportation Commission (MTC) in the Spring and Fall of 1990. The modal alternatives include drive alone, shared-ride, and transit. The departure time choice is represented by six time-periods: early morning (12:01a.m.-7a.m.), a.m. peak (7:01a.m.-9a.m.), a.m. offpeak (9:01a.m.-12noon), p.m. offpeak (12:01p.m.-3p.m.), p.m. peak (3:01p.m.-6p.m.), and evening (6:01p.m.-12 midnight). For some individual trips, modal availability is a function of time-of-day (for

example, transit mode may be available only during the a.m. and p.m. peak periods). Such temporal variations in modal availability are accommodated by defining the feasible set of joint choice alternatives for each individual trip. Level of service data were generated for each zonal pair in the study area and by five time periods: early morning, a.m. peak, mid-day, p.m. peak, and evening. The sample used in Bhat's paper comprises 3000 home-based social/recreational person-trips obtained from the overall single-day travel diary sample. The mode choice shares in the sample are as follows: drive alone (45.7%), shared-ride (51.9%) and transit (2.4%). The departure time distribution of home-based social-recreational trips in the sample is as follows: Early morning (4.6%), a.m. peak (5.5%), a.m. offpeak (10.3%), p.m. offpeak (17.2%), p.m. peak (16.1%), and evening (46.3%).

Bhat estimated four different models of mode-departure time choice: (1) the multinomial logit (MNL) model, (2) the mixed multinomial logit model which accommodates shared unobserved random utility attributes along the departure time dimension only (the MMNL-T model), (3) the mixed multinomial logit model which accommodates shared unobserved random utility attributes along the mode dimension only (the MMNL-M model), and (4) the proposed mixed multinomial logit model which accommodates shared unobserved attributes along both the dimensions of mode and departure time (the MMNL-MT model). In the MMNL models, the sensitivity among joint choice alternatives sharing the same mode (departure time) were allowed to vary across modes (departure times). It is useful to note that such a specification generates heteroscedasticity in the random error terms across the joint choice alternatives. In the MMNL-T and MMNL-MT models, the shared unobserved components specific to the morning departure times (*i.e.*, early morning, a.m. peak, and a.m. offpeak periods) were statistically insignificant. Consequently, the MMNL-T and MMNL-MT model results restricted these components to zero.

The level-of-service parameter estimates, implied money values of travel time, data fit measures, and the variance parameters in $[\Sigma]$ and $[\Omega]$ from the different models are presented in Table 4. The signs of the level-of-service parameters are consistent with *a priori* expectations in all the models. Also, as expected, travelers are more sensitive to out-of-vehicle travel time than in-vehicle travel time. A comparison of the magnitudes of the level-of-service parameter estimates across the four specifications reveals a progressively increasing magnitude as we move from the MNL model to the MMNL-MT model (this is an expected result since the variance before scaling is larger in the MNL model compared to the mixture

models, and in the MMNL-M and MMNL-T models compared to the MMNL-MT model; see Revelt and Train, 1998 for a similar result). The implied money values of in-vehicle and out-of-vehicle travel times are lesser in the mixed multinomial logit models relative to the MNL model.

The four alternative models in Table 4 can be evaluated formally using conventional likelihood ratio tests. A statistical comparison of the multinomial logit model with any of the mixture models leads to the rejection of the multinomial logit. Further likelihood ratio tests among the MMNL-M, MMNL-T, and MMNL-MT models result in the clear rejection of the hypothesis that there are shared unobserved attributes along only one dimension; that is, the tests indicate the presence of statistically significant shared unobserved components along both the mode and departure time dimensions (the likelihood ratio test statistic in the comparison of the MMNL-T model with the MMNL-MT model is 14.2; the corresponding value in the comparison of the MMNL-M model with the MMNL-MT model is 23.8; both these values are larger than the chi-squared distribution with 3 degrees of freedom at any reasonable level of significance). Thus, the MNL, MMNL-T, and MMNL-M models are mis-specified.

The variance parameters provide important insights regarding the sensitivity of joint choice alternatives sharing the same mode and departure time. The variance parameters specific to departure times (in the MMNL-T and MMNL-MT models) show statistically significant shared unobserved attributes associated with the afternoon/evening departure periods. However, as indicated earlier, there were no statistically significant shared unobserved components specific to the morning departure times (*i.e.*, early morning, a.m. peak, and a.m. offpeak periods). The implication is that home-based social-recreational trips pursued in the morning are more flexible and more easily moved to other times of the day than trips pursued later in the day. Social-recreational activities pursued later in the day may be more rigid because of scheduling considerations among household members and/or because of the inherent temporal "fixity" of late-evening activities (such as attending a concert or a social dinner). The magnitude of the departure time variance parameters reveal that late evening activities are most rigid, followed by activities pursued during the p.m. offpeak hours. The p.m. peak social-recreational activities are more flexible relative to the p.m. offpeak and late-evening activities. The variance parameters specific to the travel modes (in the MMNL-M and MMNL-MT models) confirm the presence of common unobserved attributes among joint choice alternatives that share the same mode; thus, individuals tend to maintain their current travel mode when

confronted with transportation control measures such as ridesharing incentives and auto-use disincentives. This is particularly so for individuals who rideshare, as can be observed from the higher variance associated with the shared-ride mode relative to the other two modes. In the context of home-based social-recreational trips, most ridesharing arrangements correspond to travel with children and/or other family members; it is unlikely that these ridesharing arrangements will be terminated after implementation of transportation control measures such as transit-use incentives.

The different variance structures among the four models imply different patterns of inter-alternative competition. To demonstrate the differences, Table 5 presents the disaggregate self- and cross-elasticities (for a person-trip in the sample with close-to-average modal level-of-service values) in response to peak period pricing implemented in the p.m. peak (*i.e.*, a cost increase in the "drive alone-p.m. peak" alternative). All morning time periods are grouped together in the table since the cross-elasticities for these time periods are the same for each mode (due to the absence of shared unobserved attributes specific to the morning time periods).

The MNL model exhibits the familiar Independence from Irrelevant Alternatives (IIA) property (that is, all cross-elasticities are equal). The MMNL-T model shows equal cross-elasticities for each time period across modes, a reflection of not allowing shared unobserved attributes along the modal dimension. However, there are differences across time periods for each mode. First, the shift to the shared ride-p.m. peak and transit-p.m. peak is more than to the other non-p.m. peak joint choice alternatives. This is, of course, because of the increased sensitivity among p.m.-peak joint choice alternatives generated by the error variance term specific to the p.m. peak period. Second, the shift to the evening-period alternatives are lower compared to the shift to the p.m. offpeak period alternatives for each mode. This result is related to the heteroscedasticity in the shared unobserved random components across time periods. The variance parameter in Table 4 associated with the evening period is higher than that associated with the p.m. offpeak period; consequently, there is less shift to the evening alternatives (see Bhat, 1995 for a detailed discussion of the inverse relationship between cross-elasticities and the variance of alternatives). The MMNL-M model shows, as expected, a heightened sensitivity of drive alone alternatives (relative to the shared-ride and transit alternatives) in response to a cost increase in the DA-p.m. peak alternative. The higher variance of the unobserved attributes specific to shared-ride (relative to transit; see Table 4) results in the lower

cross-elasticity of the shared-ride alternatives compared to the transit alternatives. The MMNL-MT model shows higher cross-elasticities for the drive alone alternatives as well as for the non-drive alone p.m. peak period alternatives since it allows shared-unobserved attributes along both the mode and time dimensions.

The drive-alone p.m. peak period self-elasticities in Table 5 are also quite different across the models. The self-elasticity is lower in the MMNL-T model relative to the MNL mode. The MMNL-T model recognizes the presence of temporal rigidity in social-recreational activities pursued in the p.m. peak. This is reflected in the lower self-elasticity effect of the MMNL-T model. The self-elasticity value from the MMNL-M model is larger than that from the MMNL-T model. This is because individuals are likely to maintain their current travel mode (even if it means shifting departure times) in the face of transportation control measures. But the MMNL-T model accommodates only the rigidity effect in departure time, not in travel mode. As a consequence, the rigidity in mode choice is manifested (inappropriately) in the MMNL-T model as a low drive alone p.m.-peak self-elasticity effect. Finally, the self-elasticity value from the MMNL-MT model is lower than the value from the MMNL-M models. The MMNL-M model ignores the rigidity in departure time; when this effect is included in the MMNL-MT model, the result is a depressed self-elasticity effect.

The substitution structures among the four models imply different patterns of competition among the joint mode-departure time alternatives. We now turn to the aggregate self- and cross-elasticities to examine the substantive implications of the different competition structures for the level-of-service variables.

Table 6 provides the cost elasticities obtained for the drive alone and transit joint-choices in response to a congestion pricing policy implemented in the p.m. peak. The aggregate cost elasticities reflect the same general pattern as the disaggregate elasticities discussed earlier. Some important policy-relevant observations that can be made from the aggregate elasticities are as follows. The DA-p.m. peak self-elasticities show that the MNL and MMNL-T models under-estimate the decrease in peak period congestion due to peak-period pricing, while the MMNL-M model over-estimates the decrease. Thus, using the DA-p.m. peak cost self-elasticities from the MNL and MNL-T models will make a policy analyst much more conservative than (s)he should be in pursuing peak-period pricing strategies. On the other hand, using the DA-p.m. peak cost self-elasticity from the MMNL-M model provides an overly-optimistic projection of the congestion alleviation due to peak period pricing. From a transit standpoint, the MNL and

MMNL-T under-estimate the increase in transit share across all time periods due to p.m. peak period pricing. Thus, using these models will result in lower projections of the increase in transit ridership and transit revenue due to a peak period pricing policy. The MMNL-M model under-estimates the projected increase in transit share in all the non-evening time periods, and over-estimates the increase in transit share for the evening time period. Thus, the MNL, MMNL-T, and MMNL-M models are likely to lead to inappropriate conclusions regarding the necessary changes in transit provision to complement peak-period pricing strategies.

5. CONCLUSIONS

This paper presents the structure, estimation techniques, and transport applications of three classes of discrete choice models: heteroscedastic models, GEV models, and flexible structure models. Within each class, alternative formulations are discussed. The formulations presented are quite flexible (this is especially the case with the flexible structure models), though estimation using the maximum likelihood technique requires the evaluation of one-dimensional integrals (in the heteroscedastic extreme value model) or multi-dimensional integrals (in the flexible model structures). However, these integrals can be approximated using Gaussian quadrature techniques or simulation techniques. In this regard, the recent use of quasi-Monte Carlo (QMC) simulation techniques seems to be particularly effective.

The advanced model structures presented in this chapter should not be viewed as substitutes 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 super-imposed 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 confluence of continued careful structural specification with the ability to accommodate very flexible substitution patterns/unobserved heterogeneity should facilitate the application of behaviorally rich structures in transportation-related discrete choice modeling in the years to come.

References

Ben-Akiva, M. and D. Bolduc (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working paper, Department of Civil and Environmental engineering, Massachusetts Institute of Technology, Cambridge, MA and Département d'économique, Université Laval, Sainte-Foy, Qc, Canada.

Ben-Akiva, M., and B. Francois (1983). Homogenous generalized extreme value model. Working paper, Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge.

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

Bhat, C.R. (1997). Recent methodological advances relevant to activity and travel behavior analysis. Invitational resource paper presented at the International Association of Travel Behavior Research Conference to be held in Austin, Texas, September 1997.

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

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

Bhat, C.R. (1998c). An analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research*, 32B, pp. 361-371.

Bhat, C.R. (2000) Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science*, 34, 2, 228-238.

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

Bhat, C.R. (2002) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Forthcoming, *Transportation Research*.

Bhat, C.R. and S. Castelar (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*, 36B, 593-616.

Bhat, C.R., and J. Guo (2002). A spatially correlated logit model: formulation and application to residential choice modeling. Draft working paper, Department of Civil Engineering, The University of Texas at Austin.

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

Bresnahan, T.F., Stern, S., and M. Trajtenberg (1997). Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. *RAND Journal of Economics*, 28 (0), 17-44.

Brownstone, D. (2000). Discrete choice modeling for transportation. Resource paper presented at the 2000 IATBR Conference, The Gold Coast, Australia, July.

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

Brownstone, D., Bunch, D.S., and K. Train (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research*, 34B, 315-338.

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

Chu, C. (1989). A paired combinatorial logit model for travel demand analysis, Proceedings of the Fifth World Conference on Transportation Research, 295-309, Ventura, CA.

Daganzo, C. (1979). *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Academic Press, New York.

Daganzo, C.F. and M. Kusnic (1993). Two properties of the nested logit model, *Transportation Science*, 27, 395-400.

Daly, A.J. and S. Zachary (1978). Improved multiple choice models. In D.A. Hensher and M.Q. Dalvi (eds.) *Determinants of Travel Choice*, Saxon House, Westmead.

Fang, K.-T and Y. Wang (1994). *Number-Theoretic Methods in Statistics*. Chapman and Hall, London.

Forinash, C.V., and F.S. Koppelman (1993). Application and interpretation of nested logit models of intercity mode choice. *Transportation Research Record*, 1413, 98-106.

Han, A. and S. Algers (2001). A mixed multinomial logit model for route choice behavior. Presented at the 9th WCTR Conference, Seoul, Korea, July.

Hellekalek, P. (1984). Regularities in the distribution of special sequences, *Journal of Number Theory*, 18, 41-55.

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

Hensher, D.A. (1998). Extending valuation to controlled value functions and non-uniform scaling with generalized unobserved variances. In Gärling, T., Laitila, T., and Westin, K. *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. Technical Paper, Institute of Transport Studies, The University of Sydney, Australia.

Hensher, D.A. (2000). Measurement of the valuation of travel time savings. Forthcoming, special issue of *Transportation Economics and Policy* in honor of Michael E. Beesley.

Hensher, D.A., and W. Greene (2000). Choosing between conventional, electric, and UPG/LNG vehicle in single vehicle households. Technical paper, Institute of Transport Studies, University of Sydney, Australia.

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

Johnson, N. and S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions*, John Wiley, New York, Chapter 21.

Kocis, L. and W.J. Whiten (1997). Computational investigations of low-discrepancy sequences, *ACM Transactions on Mathematical Software*, 23, 2, 266-294.

Koppelman, F.S. and V. Sethi (2000). Closed-form discrete-choice models. In D. Hensher and K. Button (eds) Handbook of Transport Modelling, Pergamon, 211-225.

Koppelman, F.S. and C-H Wen (2000). The paired combinatorial logit model: properties, estimation and application. *Transportation Research*, 34B, 75-89.

KPMG Peat Marwick and F.S. Koppelman (1990). Analysis of the market demand for high speed rail in the Quebec-Ontario corridor. Report produced for Ontario/Quebec Rapid Train Task Force. KPMG Peat Marwick, Vienna, VA.

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, Wiley, New York.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, 672, 72-77.

Mehndiratta, S. (1997). Time-of-day effects in intercity business travel. Ph.D. thesis, Department of Civil Engineering, University of California, Berkeley (1996).

Munizaga, M.A., B.G. Heydecker, and J. Ortuzar (2000). Representation of heteroscedasticity in discrete choice models. *Transportation Research*, 34B, 219-240.

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

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

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

Shaw, J.E.H. (1988). A quasirandom approach to integration in Bayesian statistics. *The Annals of Statistics*, 16, 2, 895-914.

Small, K.A. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55(2), 409-424.

Stroud, A.H. (1971) *Approximate Calculation of Multiple Integrals*. Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research*, 35B, 643-666.

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

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

Train, K. (2001). A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. Technical paper, Department of Economics, University of California, Berkeley.

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

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

Williams, H.C.W.L (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning*, 9A, 285-344.

Vovsha, P. (1997). The cross-nested logit model: application to mode choice in the Tel-Aviv metropolitan area. Presented at the 1997 Annual Transportation Research Board Meeting, Washington, D.C.

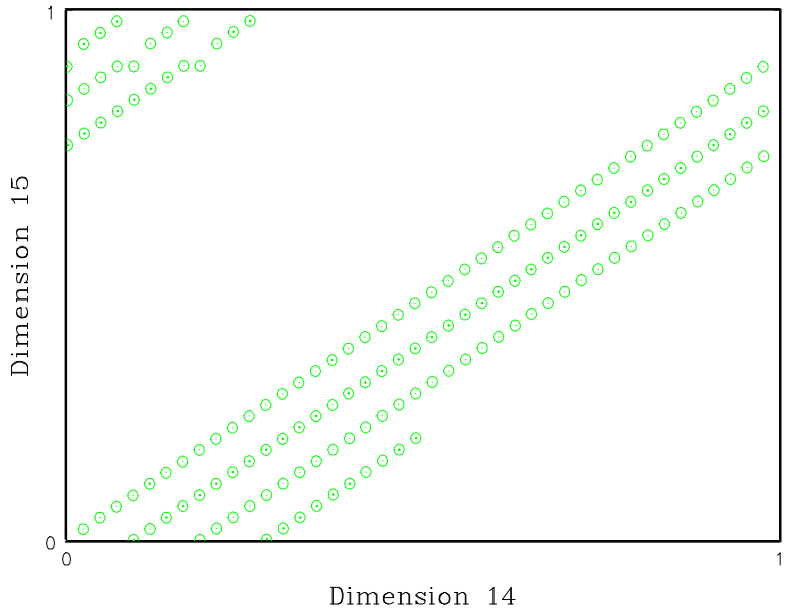
Appendix

Permutations for Scrambled Halton Sequences

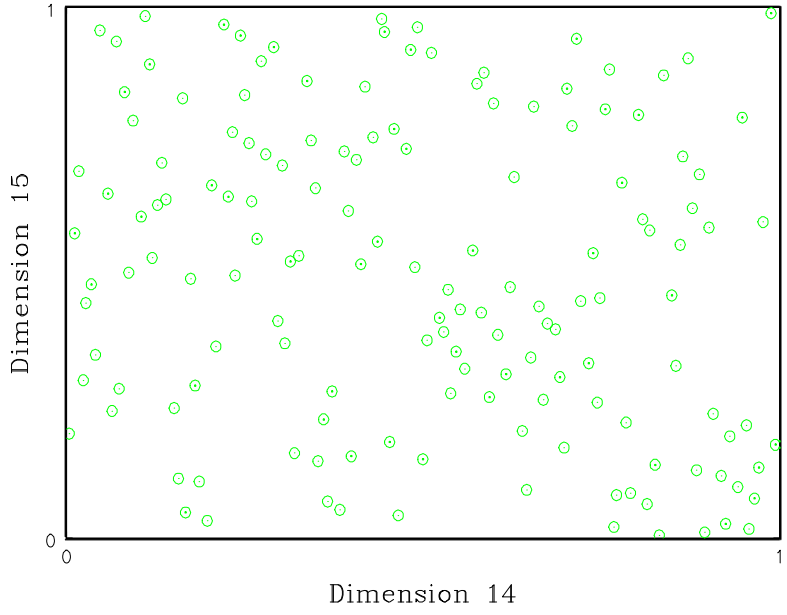
Prime *r*	Permutation of (0 1 2 ... *r*-1)
2	(0 1)
3	(0 2 1)
5	(0 3 1 4 2)
7	(0 4 2 6 1 5 3)
11	(0 5 8 2 10 3 6 1 9 7 4)
13	(0 6 10 2 8 4 12 1 9 5 11 3 7)
17	(0 8 13 3 11 5 16 1 10 7 14 4 12 2 15 6 9)
19	(0 9 14 3 17 6 11 1 15 7 12 4 18 8 2 16 10 5 13)
23	(0 11 17 4 20 7 13 2 22 9 15 5 18 1 14 10 21 6 16 3 19 8 12)
29	(0 15 7 24 11 20 2 27 9 18 4 22 13 26 5 16 10 23 1 19 28 6 14 17 3 25 12 8)

Source: Braaten and Weller (1979)

List of Figures

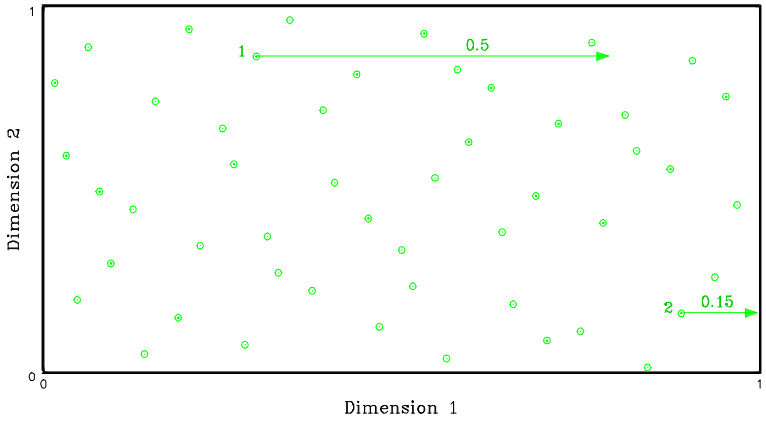


Standard Halton sequence

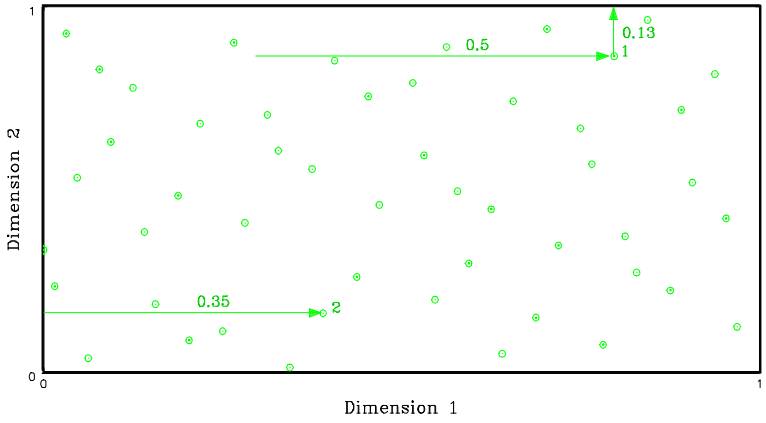


Scrambled Halton sequence

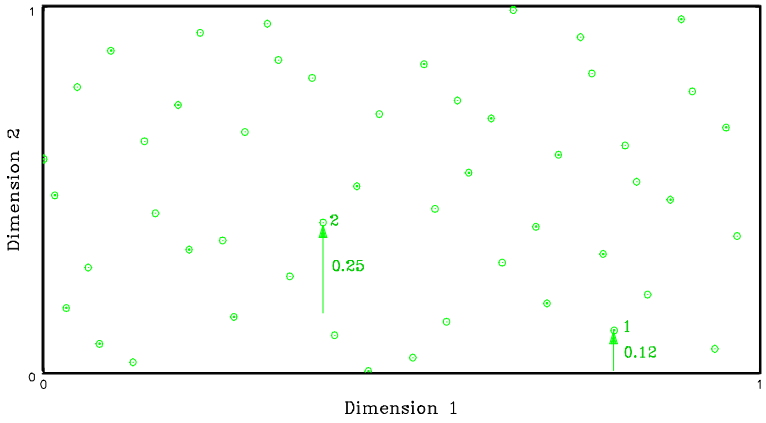
Fig.1. 150 Draws of Standard and Scrambled Halton Sequences



Standard Halton sequence



Standard Halton sequence shifted by 0.5 in Dim. 1



Standard Halton sequence shifted by 0.5 in Dim. 1 and 0.25 in Dim. 2

Fig. 2. Shifting the Standard Halton Sequence

List of Tables

Table 1. Intercity mode choice estimation results

Variable	Multinomial Logit		Nested Logit with Car and Train Grouped		Heteroscedastic Extreme Value Model	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Mode Constants (car is base)						
Train	-0.5396	-1.55	-0.6703	-2.14	-0.1763	-0.42
Air	-0.6495	-1.23	-0.5135	-1.31	-0.4883	-0.88
Large City Indicator (car is base)						
Train	1.4825	7.98	1.3250	6.13	1.9066	6.45
Air	0.9349	5.33	0.8874	5.00	0.7877	4.96
Household Income (car is base)						
Train	-0.0108	-3.33	-0.0101	-3.30	-0.0167	-3.57
Air	0.0261	7.02	0.0262	7.42	0.0223	6.02
Frequency of service	0.0846	17.18	0.0846	17.67	0.0741	10.56
Travel Cost	-0.0429	-10.51	-0.0414	-11.03	-0.0318	-5.93
Travel Time						
In-Vehicle	-0.0105	-13.57	-0.0102	-12.64	-0.0110	-9.78
Out-of-Vehicle	-0.0359	-12.18	-0.0353	-13.86	-0.0362	-8.64
logsum Parameter[4]	1.0000	-	0.9032	1.14	1.0000	-
Scale Parameters (car parm.= 1)[5]						
Train	1.0000	-	1.0000	-	1.3689	2.60
Air	1.0000	-	1.0000	-	0.6958	2.41
Log Likelihood At Convergence[6]	-1828.89		-1828.35		-1820.60	
Adjusted L'hood Ratio Index	0.3525		0.3524		0.3548	

[4]The logsum parameter is implicitly constrained to one in the multinomial logit and heteroscedastic model specifications. The t-statistic for the logsum parameter in the nested logit is with respect to a value of one.

[5]The scale parameters are implicitly constrained to one in the multinomial logit and nested logit models and explicitly constrained to one in the constrained "heteroscedastic" model. The t-statistics for the scale parameters in the heteroscedastic model are with respect to a value of one.

[6]The log likelihood value at zero is -3042.06 and the log likelihood value with only alternative specific constants and an IID error covariance matrix is -2837.12.

Table 2. Elasticity matrix in response to change in rail service for multinomial logit and heteroscedastic models

Rail Level of Service Attribute	Multinomial Logit Model			Heteroscedastic Extreme Value Model		
	Train	Air	Car	Train	Air	Car
Frequency	0.303	-0.068	-0.068	0.205	-0.053	-0.040
Cost	-1.951	0.436	0.436	-1.121	0.290	0.220
In-Vehicle Travel Time	-1.915	0.428	0.428	-1.562	0.404	0.307
Out-of-Vehicle Travel Time	-2.501	0.559	0.559	-1.952	0.504	0.384

Note: The elasticities are computed for a representative intercity business traveler in the corridor.

Table 3. Comparison between MNL and GNL model estimates

Variable	MNL model		GNL model	
	Parameter	Std error	Parameter	Std error
Mode constants				
Air	8.2380	0.429	6.2640	0.321
Train	5.4120	0.267	4.9810	0.285
Car	4.4210	0.301	5.1330	0.253
Bus (base)				
Frequency	0.0850	0.004	0.0288	0.002
Travel cost	-0.0508	0.003	-0.0173	0.002
In-vehicle time	-0.0088	0.001	-0.0031	0.0002
Out-of-vehicle time	-0.0354	0.002	-0.0110	0.001
Logsum parameters				
Train-car			0.0146	0.002
Air-car			0.2819	0.032
Train-car-air			0.0100	--
Allocation parameters				
Train-car nest				
Train			0.2717	0.033
Car			0.1057	0.012
Air-car nest				
Air			0.6061	0.040
Car			0.4179	0.046
Train-car-cir nest				
Train			0.5286	0.031
Car			0.2741	0.029
Air			0.3939	0.041
Train nest			0.1998	0.025
Car nest			0.2024	0.032
Bus nest			1.0000	
Log-likelihood at convergence	-2784.6		-2711.3	
Likelihood ratio index				
vs. zero	0.4896		0.5031	
vs. market share	0.3205		0.3382	
Value of time (per hour)				
In-vehicle time	C$ 10		C$ 11	
Out-of-vehicle time	C$ 42		C$ 38	
Significance test rejecting MNL model ($?^2$, DF, Sig.)	--		146.6, 11, <0.0001	

Source: Wen and Koppelman (2001)

Table 4. Level of service parameters, implied money values of travel time, data fit measures, and error variance parameters

Attributes/data fit measures	MNL model	MMNL-T model	MMNL-M model	MMNL-MT model
Level of service[7]				
Travel cost (in cents)	-0.0031 (-3.13)	-0.0036 (-3.02)	-0.0044 (-2.88)	-0.0045 (-2.83)
Total travel time (in mins.)	-0.0319 (-3.15)	-0.0336 (-2.87)	-0.0382 (-3.22)	-0.0408 (-3.33)
Out-of-vehicle time/distance	-0.2363 (-3.42)	-0.2429 (-4.82)	-0.2508 (-4.19)	-0.2589 (-4.26)
Implied money values of time ($/hr)				
In-vehicle travel time	6.17	5.60	5.21	5.44
Out-of-vehicle travel time[8]	13.66	12.23	10.80	11.09
LL at Convergence[9]	-6393.6	-6382.9	-6387.7	-6375.8
Error variance parameters				
$?_{pm\ offpeak}$	-	0.8911 (2.76)	-	0.9715 (2.96)
$?_{pm\ peak}$	-	0.7418 (2.83)	-	0.3944 (1.88)
$?_{evening}$	-	1.9771 (2.70)	-	1.6421 (3.02)
$?_{drive\ alone}$	-		0.6352 (1.91)	0.5891 (1.98)
$?_{shared\ ride}$	-		1.9464 (3.06)	1.9581 (3.20)
$?_{transit}$	-		0.7657 (1.73)	0.7926 (2.07)

[7]The entries in the different columns correspond to the parameter values and their t-statistics (in parenthesis).

[8]Money value of out-of-vehicle time is computed at the mean travel distance of 6.11 miles.

[9]The LL (Log-Likelihood) at equal shares is -8601.24 and the LL with only alternative specific constants and an IID error covariance matrix is -6812.07

Table 5. Disaggregate travel cost elasticities in response to a cost increase in the drive alone (DA) mode during p.m. peak

Effect on Joint Choice Alternative	MNL model	MMNL-T model	MMNL-M model	MMNL-MT model
DA-morning periods[1]	0.0072	0.0085	0.0141	0.0165
DA-p.m. offpeak	0.0072	0.0060	0.0141	0.0131
DA-p.m. peak	-0.1112	-0.0993	-0.1555	-0.1423
DA-evening	0.0072	0.0042	0.0141	0.0099
SR-morning periods[1]	0.0072	0.0085	0.0059	0.0072
SR-p.m. offpeak	0.0072	0.0060	0.0059	0.0055
SR-p.m. peak	0.0072	0.0120	0.0059	0.0079
SR-evening	0.0072	0.0042	0.0059	0.0045
TR-morning periods[1]	0.0072	0.0085	0.0119	0.0131
TR-p.m. offpeak	0.0072	0.0060	0.0119	0.0106
TR-p.m. peak	0.0072	0.0120	0.0119	0.0150
TR-evening	0.0072	0.0042	0.0119	0.0082

[1]The morning periods include early morning, a.m. peak, and a.m. off-peak. The cross-elasticities for the morning periods within each mode with respect to a p.m. peak cost increase in the drive alone mode are the same in the mixture logit models because of the absence of shared unobserved attributes specific to the morning time periods.

Table 6. Aggregate travel cost elasticities in response to a cost increase in the drive alone (DA) mode during p.m. peak

Effect on Joint Choice Alternative	MNL model	MMNL-T model	MMNL-M model	MMNL-MT model
Drive alone (DA) alternatives				
early morning	0.0146	0.0202	0.0290	0.0392
a.m. peak	0.0125	0.0166	0.0259	0.0334
a.m. offpeak	0.0121	0.0155	0.0250	0.0317
p.m. offpeak	0.0123	0.0136	0.0254	0.0265
p.m. peak	-0.1733	-0.1536	-0.2355	-0.2192
evening	0.0146	0.0088	0.0293	0.0204
Transit (TR) alternatives				
early morning	0.0197	0.0260	0.0280	0.0371
a.m. peak	0.0188	0.0237	0.0283	0.0358
a.m. offpeak	0.0163	0.0195	0.0236	0.0291
p.m. offpeak	0.0168	0.0175	0.0246	0.0251
p.m. peak	0.0218	0.0393	0.0333	0.0485
evening	0.0205	0.0120	0.0299	0.0203