Econometric Choice Formulations:
Alternative Model Structures, Estimation Techniques, and Emerging Directions

Resource paper for workshop entitled
"Econometric Models of Choice: Formulation and Estimation"
2003 IATBR Conference, Lucerne, Switzerland

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

ABSTRACT

The last six years since the Austin IATBR conference has been a very fertile period for the germination of new conceptual, theoretical, and computational developments in the field of econometric choice models. There is a sense today of absolute control over the kind of choice behavior structures one wants to specify in empirical contexts and a renewed excitement in the field. This paper reviews these recent developments and assembles a list of recent applications of advanced discrete choice models.

1. INTRODUCTION

Econometric models of choice have witnessed a literal revolution in recent years, as the ability of the analyst to incorporate and estimate realistic behavioral structures has been enhanced considerably. There are two reasons for this revolution. One is that, after a long hiatus, new model structures are being discovered and introduced within the framework of Generalized Extreme Value (GEV) models. The flexibility that such new GEV constructs offer are very valuable, especially since the resulting choice probability and likelihood functions still retain a desirable analytic closed-form structure. Second, there has been substantial progress in simulation methods to estimate likelihood functions involving analytically intractable multidimensional integrals. This has allowed analysts to estimate practically any choice model structure, without limiting the specification to mathematically convenient, and behaviorally less desirable model forms. In regard to both the points above, it is true that there have been some slow and steady advances in choice modeling techniques over the past three decades since McFadden's pioneering work in the early 1970s. But it is by no means an exaggeration to state that the last 6 years (since the Austin IATBR conference) has been one of the most fertile periods in sowing the seeds for a new way of thinking, and applying, choice models. Specifically, these past few years has seen a surge in progress, a feeling of liberation from the "bondage" of restrictive model forms, a sense of absolute control over the kind of behavioral structures one wants to specify in empirical contexts, renewed excitement in the field, and clasped anticipation of new developments on the horizon.

The purpose of this paper is to review these recent methodological advances in econometric models of choice, and identify the challenges ahead. Several points are in order before proceeding to the remainder of the paper. First, this paper assumes a reasonably high level of familiarity with discrete choice models, and so does not belabor over the basic structures of model forms such as the multinomial logit, nested logit, probit, heteroscedastic extreme value, and mixed logit models. Readers interested in the basics of these forms are encouraged to consult a number of recent references, including Bhat (2003a; 2002), Koppelman and Sethi (2000), Silliano and Ortúzar (2005), Greene and Hensher (2003), and Train (2003). The last reference, which is a book on simulation methods for discrete choice is, in particular, a comprehensive resource for readers. Second, this paper does not address data collection, survey methodology, and data imputation issues. While good econometric modelers always realize the importance of data quality, pay attention to data-related issues, and assemble the data with great care, there is only so much that can be covered here. For interested readers, several topics regarding data collection, survey methodology, and imputation considerations have been addressed in papers presented at a recent conference in South Africa (see Jones and Stopher, 2003). Additionally, Brownstone *et al.* (2003) and Steinmetz and Brownstone (2005) are good reading sources for survey nonresponse and imputation approaches. Third, we focus on econometric discrete choice models or model forms that are very similar to econometric discrete choice models. Limited dependent variable models that combine discrete choices with continuous and/or grouped decisions (including sample selection models) are not examined here. Lewbel and Linton (2002) and the references therein provide an overview of recent developments in the area of semi-parametric and non-parametric specifications in the context of limited dependent variables, and Bhat (2002) provides an overview in the context of applications in activity and travel behavior analysis.

The rest of the paper is structured as follows. The next section discusses four classes of advanced discrete choice model structures. Section 3 presents recent advances in the area of simulation techniques to estimate econometric models with analytically intractable probability expressions. Section 3 identifies a few emerging methodological directions in discrete choice modeling. Finally, Section 4 concludes the paper with a presentation of recent applications of advanced discrete choice models.

2. ADVANCED DISCRETE CHOICE MODEL STRUCTURES

This section discusses four types of advanced discrete choice model structures: (1) The GEV class of models, (2) The mixed multinomial logit (MMNL) class of models, (3) The mixed GEV (MGEV) class of models, and (4) Other mixed discrete choice models.

2.1 The GEV Class of Models

The GEV-class of models relaxes the independence from irrelevant alternatives (IIA) property of the multinomial logit model by relaxing the independence assumption between the error terms of alternatives. In other words, a generalized extreme value error structure is used to characterize the unobserved components of utility as opposed to the univariate and independent extreme value error structure used in the multinomial logit model. There are three important characteristics of all GEV models: (1) The overall variances of the alternatives (i.e., the scale of the utilities of alternatives) are assumed to be identical across alternatives, (2) The choice probability structure takes a closed-form expression, and (3) all GEV models collapse to the MNL model when the parameters generating correlation take values that reduce the correlations between *each pair* of alternatives to zero. With respect to the last point, it has to be noted that the MNL model is also a member of the GEV class, though we will reserve the use of the term "GEV class" to models that constitute generalizations of the MNL model.

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, 1998a), the ordered GEV-nested logit (OGEV-NL) model (Whelan *et al.*, 2002) and the Product Differentiation Logit (PDL) model (Breshanan *et al*., 1997). More recently, Wen and Koppelman (2001) proposed a general GEV model structure, which they referred to as the Generalized 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. Swait (2001) presents a network representation for the GenL model, which also applies to the GNL model.

Researchers, of course, are not restricted to the GEV structures identified above, and can generate new GEV model structures customized to their specific empirical situation. In fact, only a handful of possible GEV model structures appear to have been implemented, and there are likely to be several, yet undiscovered, model structures within the GEV class. For example, Karlstrom (2001) has proposed a GEV model that is quite different in form from all other GEV models derived in the past.

One impediment to the generation of new GEV models, however, is that the conditions developed by McFadden for qualification as a GEV structure are based on a generating function G, which may not map easily into a desired correlation structure. Recent work by Bierlaire (2002) and Daly and Bierlaire (2003) have the potential to remove this impediment. These two researchers propose a network-based structure to characterize the underlying correlation structure in any choice situation, and show how this network-based representation, if it satisfies some simple conditions (non-emptiness, finiteness, and being circuit-free), can immediately be translated to a model consistent with the GEV structure (this work constitutes a formal and rigorous extension of Swait's network representation for the GenL model). The value of Daly and Bierlaire's contribution is in facilitating the translation of intuitive correlation patterns into a GEV structure without the need to start from McFadden's mathematical conditions. In summary, the work of Daly and Bierlaire should allow the realization and exploitation of the true potential of the GEV structure to capture correlation patterns.

Of course, GEV models based on complex network representations, while allowing flexibility in substitution patterns, also entail the estimation of a substantial number of dissimilarity and allocation parameters. The net result is that the analyst will have to impose informed restrictions on these GEV models, customized to the application context under investigation.

An important point to note here is that GEV models are consistent with utility maximization only under rather strict (and often empirically violated) restrictions on the dissimilarity and allocation parameters (specifically, the dissimilarity and allocation parameters should be bounded between 0 and 1 for global consistency with utility maximization, and the allocation parameters for any alternative should add to 1). The origin of these restrictions can be traced back to the requirement that the variance of the joint alternatives be identical in the GEV models. Also, GEV models do not relax assumptions related to taste homogeneity in response to an attribute (such as travel time or cost in a mode choice model) due to unobserved decision-maker characteristics, and cannot be applied to panel data with temporal correlation in unobserved factors within the choices of the same decision-making agent. However, it is indeed refreshing to note the renewed interest and focus on GEV models today, since such models do offer computational tractability, provide a theoretically sound measure for benefit valuation, and can form the basis for formulating mixed models that accommodate random taste variations and temporal correlations in panel data (see Section 2.3).

2.2 The MMNL Class of Models

The MMNL class of models, like the GEV class of models, generalizes the MNL model. However, unlike the closed form of the GEV class, the MNL class involves the analytically intractable integration of the multinomial logit formula over the distribution of unobserved random parameters. It takes the structure shown below:

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta \mid \theta) d(\beta), \text{ where}$$
$$L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qj}}}. \tag{1}$$

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 structure in Equation (1) assumes a continuous distribution for $f(\beta)$. In fact, a discrete distribution can also be used. Such a discrete distribution may take one of two forms. If the entire vector β can take one of S possible values labeled $\beta_1,\ \beta_2,\ldots,\beta_s,\ldots,\beta_S$, and the probability of $\beta=\beta_s$ for individual q is π_{qs}, then the appropriate formula is:

$$P_{qi}(\theta)=\sum_s \pi_{qs} L_{qi}(\beta_s),\ \text{where}$$
$$L_{qi}(\beta_s)=\frac{e^{\beta_s' x_{qi}}}{\sum_j e^{\beta_s' x_{qj}}},\ \text{and} \sum_s \pi_{qs}=1\ \text{for all}\ q. \quad (2)$$

π_{qs} in the equation above can be further parameterized as a function of observable individual attributes using any function that satisfies $\sum_s \pi_{qs}=1$ (usually a multinomial logit form is used). In this first form of a discrete distribution for the vector β, the MMNL model becomes equivalent to the latent-class model that has been used in marketing and in transportation (see Kamakura and Russell, 1989; Greene and Hensher, 2003; Bhat, 1997; Gupta and Chintagunta, 1994). A second possible discrete distribution approach is to use a non-parametric form separately for each coefficient in the model. This approach does not impose any prior continuous distribution function, and allows the data to identify the mass points and the associated mixing weights for each coefficient separately. Of course, such a non-parametric distribution specification can lead to convergence problems unless the number of mass points for each coefficient is limited to a small number.

Andrews *et al.* (2002) compare the continuous distribution assumption with the first of the two forms of discrete distributions discussed above for a mixed logit model estimated using repeated choice data. They find that the continuous distribution performs poorly in terms of parameter recovery and performance on a validation sample when the number of choice occasions from the same decision-making agent is small (3 or less). However, with a higher number of choices per household, there are no differences in parameter recovery and predictive validity between the discrete and continuous heterogeneity representations, though the continuous representation has an advantage in data fit in the estimation sample. Their results show that both the continuous and discrete distributions are very robust to violations of the assumed distributional assumptions, and they conclude that the selection between continuous and discrete distributions for consumer heterogeneity "is a matter of opinion and personal preference". Greene and Hensher (2003) also compare the continuous distribution assumption with a discrete distribution using a 2000 stated preference survey of long distance travelers in New Zealand. They reach the same conclusions as Andrews *et al.* (2002), and emphasize the need for additional empirical investigation comparing the continuous and discrete forms.

In the rest of this section, we focus on a continuous distribution assumption for $f(\beta)$, since this has been the more dominant assumption under the label of mixed logit.

The first applications of the mixed logit structure of Equation (1) appear to have been by Boyd and Mellman (1980) and Cardell and Dunbar (1980). However, these were not individual-level models and, consequently, the integration inherent in the mixed logit formulation had to be evaluated only once for the entire market. Train (1986) and Ben-Akiva *et al.* (1993) applied the mixed logit to customer-level data, but considered only one or two random coefficients in their specifications. Thus, they were able to use quadrature techniques for estimation. The first applications to realize the full potential of mixed logit by allowing several random coefficients simultaneously include Revelt and Train (1998) and Bhat (1998b), both of which were originally completed in the early 1996 and exploited the advances in simulation methods.

The MMNL model structure of Equation (1) can be motivated from two very different (but formally equivalent) perspectives (see Bhat, 2000a). 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) or a combination of the two. Examples of the error-components motivation in the literature include Brownstone and Train (1999), Bhat (1998c), Jong *et al.* (2002a,b), Whelan *et al.* (2002), and Batley *et al.* (2001a,b). The reader is also referred to the work of Walker and her colleagues (Ben-Akiva *et al.*, 2001; Walker, 2002) and Munizaga and Alvarez-Daziano (2002) for important identification issues in the context of the error components MMNL model. Examples of the random-coefficients structure include Revelt and Train (1998), Bhat, (2000b), Hensher (2001), and Rizzi and Ortúzar (2003).

A normal distribution is assumed for the density function $f(.)$ in Equation (1) when an error-components structure forms the basis for the MMNL model. However, while a normal distribution remains the most common assumption for the density function $f(.)$ for a random-coefficients structure, other density functions may be more appropriate. For example, a log-normal distribution may be used if, from a theoretical perspective, an element of β has to take the same sign for every individual (such as a negative coefficient on the travel cost parameter in a travel mode choice model). Other distributions that have been used in the literature include triangular and uniform distributions (see Revelt and Train, 2000; Train, 2001; Hensher and Greene, 2003) and the Rayleigh distribution (Siikamaki and Layton, 2001). The triangular and uniform distributions have the nice property that they are bounded on both sides, thus precluding the possibility of very high positive or negative coefficients for some decision-makers as would be the case if normal or log-normal distributions are used. By constraining the mean and spread to be the same, the triangular and uniform distributions can also be customized to cases where all decision-makers should have the same sign for one or more coefficients. The Rayleigh distribution, like the lognormal distribution, assures the same sign of coefficients for all decision-makers.[1]

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

A few notes are in order here about the MMNL model vis-à-vis the MNP model. First, both these models are very flexible in the sense of being able to capture random taste variations and flexible substitution patterns. Second, both these models are able to capture temporal correlation over time, as would normally be the case with panel data. Third, the MMNL model is

[1] The reader is referred to Hess and Axhausen (2005) for a review of alternative distribution forms and the ability of these distributed forms to approximate several different types of true distributional forms.

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

2.3 The Mixed GEV Class of Models

The MMNL class of models is very general in structure and can accommodate both relaxations of the IID assumption as well as unobserved response homogeneity within a simple unifying framework. Consequently, the need to consider a mixed GEV class may appear unnecessary. However, there are instances when substantial computational efficiency gains may

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

be achieved using a MGEV structure. Consider, for instance, Bhat and Guo's (2004) model for household residential location choice. It is possible, if not very likely, that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow contiguous alternatives to be correlated. In the MMNL structure, such a correlation structure may be imposed through the specification of a multivariate MNP-like error structure, which will then require multidimensional integration of the order of the number of spatial units (see Bolduc *et al.*, 1996). On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation.[3] However, the GEV model structure of Bhat and Guo cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such random coefficients, leading to a parsimonious and powerful MGEV structure. Thus, in a case with 1000 spatial units (or zones), the MMNL model would entail a multidimensional integration of the order of 1000 plus the number of random coefficients, while the MGEV model involves multidimensional integration only of the order of the number of random coefficients (a reduction of dimensionality of the order of 1000!).

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

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

alternative error correlation structure represents a powerful approach to capture random taste variations and complex substitution patterns.

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

2.4 Other Mixed Discrete Choice Models

The mixing of a distribution with a closed form analytic expression has application far beyond the MMNL and MGEV structures discussed above. For example, random coefficients can be imposed in an ordered-response multinomial model or a count model. For instance, Bhat (1999) uses a mixed ordered-response model to analyze stop-making of workers during the evening commute home. He accommodates unobserved heterogeneity across individuals in the propensity to participate in evening commute stops due to variation in sensitivity to commute travel time, work duration, and work departure times.

3. SIMULATION ESTIMATION TECHNIQUES

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

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

3.1 The Monte-Carlo Method

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

3.2 The Quasi-Monte Carlo Method

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

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

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

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

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

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

3.3 The Hybrid Method

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

Bhat (2003b) describes a process to randomize QMC sequences for use in simulation estimation. This process, based on Tuffin's (1996) randomization procedures, is described intuitively and mathematically by Bhat in the context of a single multidimensional integral. We discuss the intuitive perspective here, which is illustrated in Figure 1 in two dimensions. The first diagram in Figure 1 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 1 is also well-distributed because the relative positions of the points do not change from that in the first diagram; there is simply a shift of the overall pattern of points. The last diagram in Figure 1 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 $u \in \{0,1\}^S$ 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.

3.4 Summary on Simulation Estimation of Mixed Models

The discussion above shows the substantial progress in simulation methods, and the arrival of quasi-Monte Carlo (QMC) methods as an important breakthrough in the simulation estimation of advanced discrete choice models. The discovery and application of QMC sequences for discrete choice model estimation is a watershed event and has fundamentally changed the way we think about, specify, and estimate discrete choice models. However, lest we should leave the impression that the use and application of QMC methods has matured to the point that little addition scientific enquiry is needed, it is also important to identify some quirks that have been noticed in QMC-based estimation. Specifically, it has been noticed that QMC-based methods, on occasion, do provide results that are much worse than the norm for such methods. Similarly, using fewer QMC-draws in simulation, sometimes, tend to provide substantially better results than using a higher number of QMC draws. These results are perplexing; it is unclear if these unexpected results are due to certain properties of QMC sequences that we are yet to understand or whether it is due to the optimization algorithm used. In either case, a better understanding of the cause should provide insights to further improvement in QMC-based simulation methods for discrete choice modeling.

Notwithstanding the issues raised above, it must be emphasized that QMC methods have always provided far superior results than PMC methods and with much fewer draws. There appears to be little doubt that QMC methods will become the “bread and butter” of simulation techniques in the field in the years to come.

3.5 Bayesian Estimation of the Mixed Models

Some recent papers (Brownstone, 2001; Train, 2001; Silliano and Ortúzar, 2005) have considered a Bayesian estimation approach for MMNL model estimation as opposed to the classical estimation approaches discussed above (see also Train, 2003, Chapter 12, for a complete discussion of Bayesian methods). By considering the individual-specific parameters to be parameters themselves (in addition to the population mean and variance of the distribution of these parameters) and drawing from the posterior distributions using Gibbs sampling, the Bayesian approach avoids the need for integration. However, convergence to draws from the posterior distribution requires adequate repeated iterations of draws of the various sets of parameters. The number of iterations required for this convergence is anything but straightforward to determine. The net result is that the problem of convergence in likelihood function in the classical approach is replaced with the problem of convergence to the posterior distribution in the Bayesian approach.

The general results from comparisons of the classical and Bayesian 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. In the overall, the results suggest that the choice between the two estimation approaches depends more on interpretational ease in the empirical context under study rather than computational efficiency considerations.

4. OTHER EMERGING METHODOLOGICAL ISSUES IN DISCRETE CHOICE MODELING

4.1 Endogeneity of Variables in Discrete Choice Model

In several discrete choice contexts, there is the possibility that certain "independent" variables are not truly exogenous. Rather, the value of the variable is correlated with the unobserved factors that impact the utility/preference for an alternative. We discuss a few such examples below.

One example of endogeneity is the effect of cost on recreational site selection. Consider an individual choosing among several parks in an urban area. Park attributes of importance to the individual's choice may include availability of biking and hiking paths, land-sports facilities (basketball court, sand volleyball, *etc.*), water-sports facilities, and clean/modern showering places. An analyst modeling choice of recreational park may have access to some, but not all of these park-related characteristics. Assume, for example, that the analyst does not have information on how clean/modern the showering places are at the alternative park sites. Perhaps, there are also other factors known to the consumer and to the park manager, but unobserved to the analyst, such as aspects of style and prestige associated with a park. In these instances, the park entrance fee is set by the park manager based on these unobserved (to the analyst) park characteristics. These same unobserved characteristics also enter into the utility function of the consumer, generating a "spurious" correlation in prices and preferences; a higher-priced park is preferred by a consumer due to unobserved common characteristics affecting park entrance prices and consumer preferences. If this correlation is not controlled for, the result is an undervaluation of the effect of price on recreational site choice.

A second example is a case of a household choosing between alternative television reception options, such as cable or dish. Some aspects of each of these options, such as the quality of programming, may not be available to the analyst. These aspects, however, influence the price set by cable and dish companies as well as the preferences of customers. Petrin and Train (2002) undertake such an analysis, and empirically show that the price coefficient is substantially underestimated if the endogeneity in price is not recognized.

A third example is the effect of Information and Communication Technologies (ICT) on activity-travel behavior. Consider the effect of internet shopping at home on participation in out-of-home shopping episodes. There may be common unobserved factors affecting both internet shopping at home and out-of-home participation in shopping episodes (see Bhat *et al.* 2003). For instance, it is possible that an individual who has a shopping-oriented lifestyle is more likely to internet-shop, as well as be more likely to participate in out-of-home shopping activities. If this association is ignored, the intrinsic complementarity in internet shopping and out-of-home shopping reduces the magnitude of the true substitution effect of internet shopping on out-of-home shopping participation.

In all the cases discussed above, the endogeneity of an "independent" variable in the discrete choice model leads to biased parameters unless the endogeneity is recognized (see Villas-Boas and Winer, 1999). While the problem of endogeneity is by no means a new issue in econometrics, much earlier work has been focused on linear models and not on non-linear models. Berry *et al.* (1995), Goolsbee and Petrin (2002), Blundell and Powell (2001), Villas-Boas and Winer (1999), Petrin and Train (2002), and Bhat *et al.* (2003) have recently provided methods to account for endogeneity in general non-linear models and discrete choice models. The most commonly used approach is to write the endogenous "independent" variable as a function of instrument variables (which can, of course, include other independent variables) and an error term. This error term is allowed to correlate with the error term in the discrete choice model, thus absorbing the part of unobserved utility that is correlated with the endogenous "independent" variable. Then, the remaining part of utility is not correlated with the endogenous "independent", thus allowing consistent estimation of the effect of the endogenous variable. If there are multiple endogenous variables, the same technique is followed for each variable. Estimation will, in general, involve analytically intractable integration, which can be easily achieved using simulation techniques.

We provide an example of the formulation and estimation of such a model with endogenous "independent" variables in the context of the effect of two ICT-use variables (mobile phone use and computer use) on number of shopping episodes. The equation comprises three equations: one equation each for the mobile phone and computer use choices, and a third ordered-response equation for the number of shopping episodes. The equation system is presented below:

$$m_q^* = \theta' h_q + \zeta_q + \upsilon_q, m_q = 1 \text{ if } m_q^* > 0, m_q = 0 \text{ if } m_q^* \leq 0$$

$$p_q^* = \mu' r_q + \xi_q + \omega_q, p_q = 1 \text{ if } p_q^* > 0, p_q = 0 \text{ if } p_q^* \leq 0 \tag{3}$$

$$s_q^* = \delta' w_q + \beta_q' x_q \pm \zeta_q \pm \xi_q + \varepsilon_q, \quad s_q = k \text{ if } \psi_{k-1} < s_q^* < \psi_k, \quad x_q = [m_q, p_q]',$$

where q is an index for individuals, m_q^* and p_q^* are latent propensities to use mobile telephones and computers, respectively, and m_q and p_q are dummy variables representing whether or not an individual uses mobile phones and computers, respectively. h_q and r_q are column vectors of exogenous variables affecting mobile telephone use and computer use, and θ and μ are corresponding column vectors to be estimated. υ_q and ω_q are standard normal variables with a correlation ρ. This correlation term captures common unobserved factors that affect the propensity to use mobile telephones and a personal computer at home. ζ_q is a normal random error term that captures common unobserved factors influencing mobile phone use propensity and the number of shopping episodes $[\zeta_q \sim N(0, \sigma_\zeta^2)]$. This term causes "spurious" dependence in mobile phone use and the number of shopping episodes. The '±' sign in front of ζ_q in the shopping episode equation indicates that the correlation in unobserved factors between mobile phone use and shopping episodes may be positive or negative. If the sign is '+', it implies that individuals who use mobile phones are also intrinsically more likely to participate in shopping episodes. If the sign is '–', it implies that individuals who use mobile phones are intrinsically less likely to undertake shopping episodes. Of course, if such correlations are ignored, they "corrupt" the "true" dependence of the intershopping hazard on mobile phone use. This issue is discussed in more detail in the empirical results section. ξ_q is a normal random term that similarly captures common unobserved factors influencing personal computer use propensity and the number of shopping episodes; $\xi_q \sim N(0, \sigma_\xi^2)$. s_q^* is the propensity to participate in shopping episodes. w_q is a vector of individual-related attributes affecting shopping episode participation propensity and δ is a corresponding coefficient vector. x_q is a vector of ICT-use variables and β_q is a corresponding vector of individual-specific ICT-use coefficients. One can allow β_q to be a function of observed and unobserved individual characteristics by specifying the ICT use coefficient β_{ql} (l = 1,2) as a function of an observed vector y_{ql} of individual attributes and an unobserved individual-specific term η_{ql} that is assumed to be a realization from a normal distribution $\eta_{ql} \sim N(0, \sigma_{\eta l}^2)$; that is, $\beta_{ql} = \vartheta_l + \gamma_l' y_{ql} + \eta_{ql}$. ε_q is an idiosyncratic random term assumed to be standard logistically distributed.

The parameters to be estimated in the model include the θ and μ vectors, the δ, ϑ, and γ vectors in the duration model, the ψ thresholds in the shopping episode model, the ρ correlation parameter capturing the effect of common unobserved factors that affect the propensity to use mobile telephones and computers at home, the scalar variance terms σ_ζ^2 and σ_ξ^2, and the vector variance term σ_η^2. Let Ω represent a vector that includes all these parameters to be estimated, and let $\Omega_{-\sigma}$ represent a vector of all parameters except the variance terms.

Define $g_q = 2m_q - 1$ and $n_q = 2p_q - 1$. Then the likelihood function for a given value of $\Omega_{-\sigma}$ and the error terms ζ_q, ξ_q, and η_q may be written as:

$$L_q(\Omega_{-\sigma}) \mid \zeta_q, \xi_q, \eta_q = \Phi_2\left[g_q \cdot (\theta' h_q + \zeta_q), n_q \cdot (\mu' r_q + \xi_q), g_q n_q \rho\right] \times \left[G_{qt} - G_{qt-1}\right], \qquad (4)$$

where t is the actual number of shopping episodes of individual q, $\Phi_2(.)$ is the bivariate cumulative standard normal distribution, and

$$G_{qt} = L\left\{\psi_t - \left[\delta' w_q + \sum_l (\vartheta_l + \gamma_l' y_{ql} + \eta_{ql}) x_{ql} \pm \zeta_q \pm \xi_q\right]\right\}. \qquad (5)$$

$L(.)$ in the above equation is the standard cumulative logistic distribution. Next, define the following standard normal variables: $f_{q\zeta} = \dfrac{\zeta_q}{\sigma_\zeta}$, $f_{q\xi} = \dfrac{\xi_q}{\sigma_\xi}$, and $f_{q\eta l} = \dfrac{\eta_{ql}}{\sigma_{\eta l}}$ (l = 1, 2; the range of l corresponds to the number of ICTs). Also, define $f_{q\eta} = (f_{q\eta 1}, f_{q\eta 2})'$. Then the likelihood function for a given value of the parameter vector Ω and for an individual q can be written conditional on $f_{q\zeta}$, $f_{q\xi}$, and the $f_{q\eta}$ random terms as:

$$L_q(\Omega) \mid f_{q\zeta}, f_{q\xi}, f_{q\eta} = \left\{\Phi_2\left[g_q \cdot (\theta' h_q + f_{q\zeta}\sigma_\zeta), n_q \cdot (\mu' r_q + f_{q\xi}\sigma_\xi), g_q n_q \rho\right] \cdot \left[G_{(qt-1)} - G_{qt-1}\right]\right\}, \quad (6)$$

where $G_{qt} = L\left\{\psi_t - \left[\delta' w_q + \sum_l (\vartheta_l + \gamma_l' y_{ql} + f_{q\eta l}\sigma_{\eta l}) x_{ql} \pm f_{q\zeta}\sigma_\zeta \pm f_{q\xi}\sigma_\xi\right]\right\}$. The unconditional likelihood for individual q may finally be written as:

$$L_q(\Omega) = \int_{f_{q\zeta}=-\infty}^{+\infty} \int_{f_{q\xi}=-\infty}^{\infty} \int_{f_{q\eta}=-\infty}^{\infty} [L_q(\Omega) \mid f_{q\zeta}, f_{q\xi}, f_{q\eta}] \cdot d\Phi(f_{q\zeta}) d\Phi(f_{q\xi}) d\Phi(f_{q\eta}). \qquad (7)$$

The log-likelihood function is $L(\Omega) = \sum_q \ln L_q(\Omega)$, which can be maximized using simulation techniques.

4.2 Mixed RP/SP Choice Models

Stated preference (SP) and revealed preference (RP) data each have their own advantages and limitations with respect to estimation of behavioral parameters of interest (Ben Akiva *et al.*, 1992; Hensher *et al.*, 1999). This realization has led to the now long history of using both kinds of data simultaneously to analyze consumer behavior (*e.g.*, Gunn *et al.*, 1992; Ben-Akiva and Morikawa, 1990; Koppelman *et al.*, 1993; Swait and Louviere, 1993; Hensher *et al.*, 1999). However, until recently, the combination of RP and SP data has focused primarily on scaling effects, and less on other important econometric issues. Recent advances in simulation techniques have made it possible to consider several econometric issues jointly in RP/SP modeling. Specifically, four important issues need to be recognized in joint RP-SP estimation: (a) inter-alternative error structure, (b) scale difference between the RP and SP data generating processes, (c) unobserved heterogeneity effects, and (d) state dependence effects and heterogeneity in the state dependence. Each of these is discussed in turn in the subsequent paragraphs, followed by the need to consider all of the issues simultaneously within a unified RP-SP modeling framework.

The literature on joint RP-SP methods has, with few exceptions, assumed a MNL structure for the RP and SP choice processes. However, with recent methodological advances, RP-SP methods can be quite easily extended to accommodate flexible competitive patterns. Recent studies in the joint RP-SP literature that accommodate non-IID inter-alternative error structures include Cherchi and Ortúzar (2002), Hensher *et al.* (1999), and Brownstone *et al.* (2000). The first study uses a nested logit structure to accommodate correlation in public transit options. The second study accommodates heteroscedasticity across alternatives within the framework of Generalized Extreme Value (GEV) models, and the third accommodates both heteroscedasticity and correlation across alternatives within the framework of a mixed multinomial logit model.

The second econometric issue in joint RP/SP modeling is that RP and SP choices are made under different circumstances; RP choices are revealed choices in the real world, while SP choices are stated choices made in an experimental and hypothetical setting. In both the real world and experimental settings, the analyst does not have information on all the factors that influence an individual's choice. Since the RP and SP choice settings are quite different, there is no reason to believe that the variance of the unobserved factors in the RP setting will be identical to that of the variance of unobserved factors in the SP setting (see Ben-Akiva and Morikawa, 1990). There is also no *a priori* theoretical basis to suggest whether the RP error term or the SP error term will have the larger variance; this may be closely tied to the empirical context under examination. The scale difference between the RP and SP choice contexts has been recognized and accommodated in almost all previous joint RP-SP analyses.

The third econometric issue is associated with unobserved heterogeneity effects or unobserved (to the analyst) differences across decision-makers in the intrinsic preference for a choice alternative (preference heterogeneity) and/or in the sensitivity to characteristics of the choice alternatives (response heterogeneity). Stated preference methods usually involve experimental settings in which each of a sample of individuals is exposed to different stimuli corresponding to different combinations of values for the set of explanatory variables under study. It is at least possible (if not very likely) that the responses from the same individual to the different stimuli will be affected by common unobserved attributes of the individual. Of course, unobserved heterogeneity effects are not confined to the SP choice responses. The same unobserved individual-specific attributes influencing the SP choices made by an individual will also affect the RP choice of the individual. These unobserved attributes generate a correlation in utility for an alternative across all choice occasions (RP and SP choices) of the individual. The unobserved heterogeneity effects also lead (indirectly) to non-IID error structures across alternatives at each choice occasion, so that the IIA property does not hold at any choice occasion. Most RP-SP studies in the literature disregard unobserved heterogeneity. However, Morikawa (1994) accommodates unobserved preference heterogeneity in his analysis by considering an error-components structure for the RP and SP error terms. Hensher and Greene (2000) have accommodated unobserved response heterogeneity, along with inter-alternative correlation, in a study on vehicle type choice decisions.

The fourth econometric issue in joint RP-SP estimation is the state dependence effect, which refers to the influence of the actual (revealed) choice on the stated choices of the individual (the term "state dependence" is used more broadly here than its typical use in the econometrics field, where the term is reserved specifically for the effect of actual past choices on actual current choices). State dependence could manifest itself as a positive or negative effect of the choice of an alternative on the utility associated with that alternative in the stated responses. Further, in most choice situations, it is possible that the effect of state dependence is positive for some individuals and negative for others (see Ailawadi *et al.*, 1999). Besides, even within the group of individuals for which the effect is positive (or negative), the extent of the inertial (or variety-seeking) impact on stated choices may vary. Thus, joint RP-SP estimations should not only recognize state dependence, but also accommodate heterogeneity in the state dependence effect. Most RP-SP studies in transportation disregard state dependence. Axhausen *et al.*, (2004) consider the state dependence effects, but do not consider heterogeneity in the state dependence effects. Bhat and Castelar (2002) accommodate such unobserved heterogeneity in the state dependence effect of the RP choice on SP choices. Brownstone *et al.*, 1996, on the other hand, accommodate observed heterogeneity in the state dependence effect by interacting the RP choice dummy variable with sociodemographic attributes of the individual and SP choice attributes).

The fundamental reason for considering all the four modeling issues discussed above simultaneously is that there is likely to be interactions among them. Thus, accommodating restrictive inter-alternative error structures rather than flexible error structures can lead to misleading behavioral conclusions about taste effects and scaling effects in joint RP-SP models. For example, Hensher *et al.* (1999) find in their empirical analysis of freight carrier choice of firms that failure to accommodate heteroscedasticity across alternatives within each data source can lead to misleading inferences about taste and scale differences across data sources. They emphasize the need to accommodate general patterns of the error variance-covariance structure across alternatives within each data source before estimating joint RP-SP models. Louviere *et al.* (1999) also highlight this point in their review of methods to combine sources of preference data. Similarly, adopting restrictive inter-alternative structures can overstate unobserved heterogeneity in a model, and ignoring unobserved heterogeneity can overstate inter-alternative error correlations. It is also imperative that unobserved heterogeneity be incorporated in a model with state dependence (see Heckman, 1981; Keane, 1997). In the context of joint RP-SP estimation, if unobserved heterogeneity exists and the analyst ignores it, the unobserved heterogeneity can manifest itself in the form of spurious state dependence; that is, the effect of the RP choice on SP choices may be artificially overstated.[4] Similarly, if the RP choice affects SP choices and the analyst ignores this state dependence, the state dependence will manifest itself in the form of unobserved heterogeneity and overstate the level of unobserved heterogeneity. In addition,

[4] Econometrically speaking, the RP choice variable is correlated with the error term in the SP choice equation in the presence of unobserved heterogeneity. This issue is similar to the initial conditions problem in the panel data literature (Chamberlain, 1980; Degeratu, 1999).

ignoring state dependence or unobserved heterogeneity can, and generally will, lead to a bias in the effect of other coefficients in the model (Heckman, 1981; Hsiao, 1986).

The discussions above are not simply esoteric econometric considerations. For instance, unobserved heterogeneity and state dependence can have quite different policy implications, and disentangling these two effects can contribute to informed policy decisions. As an example, consider the introduction of a light rail transit service in an urban area. If RP-SP studies exploring the potential use of the light rail service suggest the presence of state dependence, but no heterogeneity, then a policy that promotes the use of light rail in the initial stages of the service might help in the long term. Policymakers might therefore want to consider a blanket subsidized fare for the first month to attract people to the new service. Then, because of state dependence effects, some of the switchers will continue to use light rail even after the period of subsidized fare expires. On the other hand, if the RP-SP study indicates no state dependence and only unobserved heterogeneity, then the blanket subsidized fare will attract riders only when the lower fare is in effect. In such a situation, it would be more useful to systematically analyze the heterogeneity effects to identify population groups that are pre-disposed to using light rail and to target them specifically for information/marketing campaigns. Of course, it is quite likely that both state dependence and unobserved heterogeneity effects will exist; the magnitude of these effects can then be used to inform the design a multi-pronged marketing strategy to attract and sustain ridership over a long-term horizon.

Bhat and Castelar (2002) propose a unified RP-SP framework that adopts a mixed multinomial logit formulation to accommodate all of the four modeling considerations discussed above.

4.3 Hazard-Based Duration Models

Hazard-based duration models are based on the concept of conditional probability of termination of duration, which recognizes the dynamics of duration; that is, it recognizes that the likelihood of ending the duration depends on the length of elapsed time since start of the duration. Hazard models provide a methodologically appropriate, intuitive, and conceptual framework to analyze duration data. It so happens that a particularly appealing and flexible form of the hazard duration model takes a discrete choice form, an observation that can be exploited to estimate increasingly advanced duration models based on recent advances in mixed-logit simulators. We discuss this issue next.

Let T_{qi} represent the continuous duration time of the i^{th} duration spell of individual q (the spell could be the duration of an episode of a particular activity purpose, the duration of the time between successive participations in a particular activity purpose, *etc.*). Let τ represent some specified time on the continuous time scale. Let $\lambda_{qi}(\tau)$ represent the hazard at continuous time τ for the i^{th} duration spell of individual q; *i.e.*, $\lambda_{qi}(\tau)$ is the instantaneous conditional probability that individual q's $(i+1)^{th}$ spell will occur at continuous time τ after her/his i^{th} participation, given that the episode does not occur before time τ:

$$\lambda_{qi}(\tau) = \lim_{\Delta \to 0^+} \frac{P(\tau < T_{qi} < \tau + \Delta \mid T_{qi} > \tau)}{\Delta} \tag{8}$$

Next, we relate the hazard rate, $\lambda_{qi}(\tau)$, to a baseline hazard rate, $\lambda_0(\tau)$, a scalar α_q capturing unobserved attributes of individual q, a vector of covariates, x_q (not including a constant), and a spell-specific unobserved component ϖ_{qi} (ϖ_{qi} corresponds to random noise across different duration spells). We accomplish this by using a proportional hazard formulation as follows:

$$\lambda_{qi}(\tau) = \lambda_0(\tau)\exp(-\alpha_q - \beta_q' x_q + \varpi_{qi}), \tag{9}$$

where β_q is a vector of individual specific coefficients. While there is no specific reason to assume any prior distribution for ϖ_{qi}, a gamma distribution is convenient for exp(ϖ_{qi}) for the reasons that will become clear later (however, a non-parametric discrete distribution or some other continuous distribution may also be assumed). The exponential specification in Equation (9) guarantees the positivity of the hazard function without placing constraints on the sign of α_q and the elements of the vector β_q .

The proportional hazard formulation of Equation (9) can be written in the following equivalent form:

$$s_{qi}^* = \ln \int_{\tau=0}^{T_{qi}} \lambda_0(\tau) d\tau = \alpha_q + \beta_q' x_q - \varpi_{qi} + \varepsilon_{qi}, \tag{10}$$

where ε_{qi} is a random term with a standard extreme value distribution: Prob($\varepsilon_{qi} < z$) = $F_\varepsilon(z)$ = 1-exp[-exp(z)].

The shape of the baseline hazard function, $\lambda_0(\tau)$ in Equation (10) has important implications for duration dynamics. One may adopt a parametric shape or a non-parametric shape for the baseline hazard. A problem with the parametric approach is that it will, in general, inconsistently estimate the hazard function when the assumed parametric form is incorrect (Meyer, 1990). The advantage of using a non-parametric form is that, even when a particular parametric form is appropriate, the resulting estimates are consistent and the loss of efficiency (resulting from disregarding information about the hazard's distribution) may not be substantial (Meyer, 1987). In the non-parametric approach, the duration scale is split into several smaller grouped discrete or intervals. Assuming a constant hazard (*i.e.*, an exponential duration distribution) within each discrete interval, one can estimate the continuous-time step-function hazard shape. The reader will note that this grouping of the time scale is not inconsistent with an underlying continuous process for the duration data. In fact, the grouping may be motivated from considerations of "rounding off" in the reporting of underlying continuous duration times and the need for accounting for the resulting tied nature of departure time data.

Let t_{qi} represent the i^{th} duration of individual q and let k be an index for the discrete intervals (thus, t_{qi} can take one of the values 1, 2, …, k, …, K). Defining τ^k as the continuous time representing the upper bound of the k^{th} discrete interval, we can write:

$$s_{qi}^* = \ln \int_{\tau=0}^{T_{qi}} \lambda_0(\tau) d\tau = \alpha_q + \beta_q' x_q - \varpi_{qi} + \varepsilon_{qi},\ t_{qi} = k \ \text{ if } \ \psi_{k-1} < s_{qi}^* < \psi_k, \psi_k = \ln \int_{\tau=0}^{\tau^k} \lambda_0(\tau) d\tau \quad (11)$$

A number of different specifications may be used for the coefficient vectors α_q and β_q in Equations (9) and (10). The simplest specification is $\alpha_q = 0$ and $\beta_q = 0$ for all individuals, and $\varpi_{qi} = 0$ for all intershopping duration spells. This, of course, corresponds to the Kaplan-Meier sample hazard. A second specification is to write $\alpha_q = 0$, but to allow heterogeneity across individuals in the effect of covariates on the hazard due to observed individual characteristics by specifying the coefficients β_{ql} (l = 1, 2, …, L) as a function of an observed vector y_{ql} of individual attributes: $\beta_{ql} = \vartheta_l + \gamma_l' y_{ql}$. The spell-specific error term ϖ_{qi} is included in this formulation. The variance of ϖ_{qi} captures the level of heterogeneity in intershopping hazard across all spells and individuals. We will refer to this specification as the deterministic coefficients duration (DCD) model. A third specification superimposes random (unobserved) individual heterogeneity over the deterministic (observed) heterogeneity of the DCD model: $\alpha_q = \delta' w_q + v_q$ and $\beta_{ql} = \vartheta_l + \gamma_l' y_{ql} + \eta_{ql}$, where v_q and η_{ql} are assumed to be normally distributed across individuals; $[v_q \sim N(0,\sigma_\nu^2); \eta_{ql} \sim N(0,\sigma_{\eta l}^2)]$. In addition, we assume that v_q is independent of each η_{ql} random term (l = 1, 2, …, L) and that the η_{ql} terms are independent of each other; v_q and η_{ql} represent individual-specific unobserved factors associated with overall preferences and effect of individual-associated attributes, respectively; the variance of ϖ_{qi} in this third specification captures within-individual heterogeneity in the hazard. We will refer to the random specification above as the random coefficients duration (RCD) model.

The parameters to be estimated in the RCD model structure include ϑ and γ vectors in the duration model, the ψ thresholds in the duration model that provide information regarding the baseline intershopping hazard profile, the scalar term σ_v^2 and the vector variance term σ_η^2. Let Ω represent a vector that includes all these parameters to be estimated, and let $\Omega_{-\sigma}$ represent a vector of all parameters except the variance terms. Then the likelihood function for a given value of $\Omega_{-\sigma}$ and the error terms η_q, ν_q, and ϖ_{qi} may be written for an individual q's i^{th} duration spell as:

$$L_{qi}(\Omega_{-\sigma}) \mid \eta_q, \nu_q, \varpi_{qi} = \Big[\Big(\exp\{-B_{(t_{qi}-1)}\exp(\varpi_{qi})\}\Big) - \Big(\exp\{-B_{t_{qi}}\exp(\varpi_{qi})\}\Big)\Big], \tag{12}$$

where t_{qi} is the actual duration of individual q in the i^{th} spell, and

$$B_{t_{qi}} = \exp\left\{\psi_{t_{qi}} - \left[\nu_q + \sum_l (\vartheta_l + \gamma_l' y_{ql} + \eta_{ql}) x_{ql}\right]\right\}. \tag{13}$$

Assuming that $c_{qi}[=\exp(\varpi_{qi})]$ is distributed as a gamma random variable with a mean one (a normalization) and variance σ_c, the likelihood function for individual q's i^{th} duration spell, unconditional on ϖ_{qi}, may be written as:

$$L_{qi}(\Omega_{-\sigma}) \mid \eta_q, \nu_q = \left\{ \int_0^{\infty} \left[\exp\{-B_{t_{qi-1}} \cdot c_{qi}\} - \exp\{-B_{t_{qi}} \cdot c_{qi}\} \right] f(c_{qi}) dc_{qi} \right\} \tag{14}$$

Using the moment-generating function properties of the gamma distribution (see Johnson and Kotz, 1970), the expression above reduces to:

$$L_{qi}(\Omega_{-\sigma}) \mid \eta_q, \nu_q = \left[1 + \sigma_c^2 B_{t_{qi-1}}\right]^{-\sigma_c^{-2}} - \left[1 + \sigma_c^2 B_{t_{qi}}\right]^{-\sigma_c^{-2}} \quad . \tag{15}$$

The gamma distribution for c_{qi} is convenient because it results in a closed-form expression in Equation (14). Next, define the following standard normal variables: $f_{q\xi} = \xi_q / \sigma_\xi$ and $f_{q\eta l} = \eta_{ql} / \sigma_{\eta l}$ (l =1, 2,…,L). Also, define $f_{q\eta} = (f_{q\eta 1}, f_{q\eta 2}, f_{q3}, ..., f_{ql}, ..., f_{qL})'$. Then the likelihood function for a given value of the parameter vector Ω and for an individual q with I_q intershopping duration spells can be written conditional on $f_{q\nu}$, and the $f_{q\eta}$ random terms as:

$$L_q(\Omega) \mid f_{q\nu}, f_{q\eta} = \prod_{i=1}^{I_q} \left\{ \left[1 + \sigma_c^2 B_{t_{qi-1}}\right]^{-\sigma_c^{-2}} - \left[1 + \sigma_c^2 B_{t_{qi}}\right]^{-\sigma_c^{-2}} \right\} \tag{16}$$

and $B_{t_{qi}} = \exp\left\{ \psi_{t_{qi}} - \left[f_{q\nu}\sigma_\nu + \sum_l (\rho_l + \gamma_l' y_{ql} + f_{q\eta l}\sigma_{\eta l})\, x_{ql} \right] \right\}$

The unconditional likelihood for individual q with I_q intershopping durations may finally be written as:

$$L_q(\Omega) = \int_{f_{q\nu}=-\infty}^{\infty} \int_{f_{q\eta}=-\infty}^{\infty} [L_q(\Omega) \mid f_{q\nu}, f_{q\eta}] \cdot d\Phi(f_{q\nu}) d\Phi(f_{q\eta}) \,. \tag{17}$$

The log-likelihood function is $L(\Omega) = \sum_q \ln L_q(\Omega)$, which can be evaluated using the mixed logit class of simulators.

5.0 APPLICATIONS OF ADVANCED DISCRETE CHOICE MODELS AND CONCLUSIONS

There have been several applications of advanced discrete choice models in the past few years. Table 1 presents various studies within the past five years, organized by model type. The model types are: Generalized Extreme Value (GEV) models, Mixed Multinomial Logit (MMNL) models, and mixed GEV models and other mixed models.

Several important observations may be made based on Table 1 and our earlier discussions. First, there have been more applications using the error-components formulation of the MMNL structure than the GEV structure. This is, at least in part, because the MMNL structure is conceptually easier to understand than the GEV structure. Further, the MMNL estimation code does not materially change because of differing covariance patterns. However, both these considerations that have favored the use of the MMNL structure are not likely to continue to be significant because of the introduction of an intuitive network based representation for GEV models and the ability to write general GEV code (restrictions on this general GEV code would provide a suite of restrictive GEV models). Second, in cases where both a GEV structure and an MMNL structure can closely capture a desired competition structure, it would seem preferable to use the GEV structure. This is because the estimation of closed form analytic structures is always more accurate than the simulation evaluation of analytically intractable structures. Third, the GEV structure cannot be applied with panel data to capture temporal correlations, or with random coefficients to accommodate unobserved taste variations. Further, the GEV structure that may closely represent a desired correlation pattern may fail empirically because of violations of the bounds of the dissimilarity and allocation parameters. Clearly, from these perspectives, the MMNL is more general than the GEV structure. Fourth, the analyst will face situations when the GEV structure alone is not adequate (for example, differential sensitivities across alternatives and random taste variations across individuals). The analyst then may have a choice of using the MMNL structure or the MGEV structure. The choice of the choice structure here would depend on the situation at hand. In some situations, the number of error components in the MMNL structure to generate the basic correlation across alternatives may be very small, in which case it may be easier for the analyst to use a single overarching MMNL structure to accommodate both the basic competition structure across alternatives and factors such as random coefficients and temporal correlations. In other cases, however, the error-components to generate the basic correlation structure across alternatives may be extremely large. If an appropriate GEV structure is available to capture this correlation structure, it would be more than worth the effort to use that GEV structure and superimpose a mixing distribution to accommodate random coefficients or temporal correlations. Such a MGEV structure, in fact, may be the only practical solution in some situations (for example, see Bhat and Guo, 2004). Fifth, Table 1 indicates that, while the number of applications of advanced discrete choice models in the area of travel behavior modeling has risen considerably in the past few years, only a small group of researchers have been involved with such methods. Hopefully, the important progress in both conceptual and computational issues in the recent past will galvanize the adoption of these rich techniques in the years to come.

Two final points before concluding. One is that the field of discrete choice has seen a quantum jump in recent years. There is a sense today of absolute control over the behavioral structures one wants to estimate in empirical contexts and renewed excitement in the field, thanks to recent conceptual and simulation developments. Second, analysts need to be careful not to get carried away with these new developments in choice modeling and focus less attention on careful model specification. The fundamental idea of discrete choice models will always continue to be the identification of systematic variations in the population. The advanced methods presented in this paper should be viewed as formulations that recognize the inevitable presence of unobserved heterogeneity across individuals and/or interactions among unobserved components affecting the utility of alternatives even after adopting the best systematic specifications there can be. In fact, a valuable contribution of recent developments in the field is precisely that they enable the confluence of careful structural specification with the ability to accommodate flexible substitution patterns and unobserved heterogeneity profiles.

ACKNOWLEDEMENTS

The author would like to thank the Southwest University Transportation Center (SWUTC) for partial funding for writing this paper and for presenting this work at the IATBR conference. Lisa Macias helped with typesetting and formatting this document.

REFERENCES

Ailawadi, K.L., K. Gedenk, and S.A. Neslin (1999) Heterogeneity and purchase event feedback in choice models: An empirical analysis with implications for model buildings, *International Journal of Research in Marketing*, 16 (3) 177-198.

Andrews, R.L., A. Ainslie, and I.S. Currim (2002) An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity, *Journal of Marketing Research*, 39 (4) 479-487.

Axhausen, K.W., A. König, G. Abay, J.J. Bates and M. Bierlaire (2004) Swiss value of travel time savings, paper presented at the 2004 European Transport Conference, Strasbourg, October 2004.

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

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

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

Ben-Akiva, M. and B. Francois (1983) Mu-homogenous generalized extreme value model, Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M. and T. Morikawa (1990) Estimation of switching models from revealed preferences and stated intentions, *Transportation Research Part A*, 24 (6) 485-495.

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

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

Ben-Akiva, M., T. Morikawa, and F. Shiroishi (1992) Analysis of the reliability of preference ranking data, *Journal of Business Research*, 24 (2) 149-164.

Berry, S., J. Levinsohn, and A. Pakes (1995) Automobile prices in market equilibrium, *Econometrica*, 63 (4) 841-890.

Bhat, C.R. (1997) Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel, *Transportation Research Part B*, 31 (1) 11-21.

Bhat, C.R. (1998a) An analysis of travel mode and departure time choice for urban shopping trips, *Transportation Research Part B*, 32 (6) 361-371.

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

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

Bhat, C.R. (1999) An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey, *Transportation Research Part B*, 33 (7) 495-510.

Bhat, C.R. (2000a) Flexible model structures for discrete choice analysis, in D.A. Hensher and K.J. Button (eds.) *Handbook of Transport Modelling*, 71-90, Pergamon, New York.

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

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

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

Bhat, C.R. (2002) Recent methodological advances relevant to activity and travel behavior analysis, in H.S. Mahmassani (ed.) *In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*, 381-414, Pergamon, Amsterdam.

Bhat, C.R. (2003a) Random utility-based discrete choice models for travel demand analysis, in K. Goulias (ed.) *Transportation Systems Planning: Methods and Applications*, 10, 1-30, CRC Press, Boca Raton.

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

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 Part B*, 36 (7) 577–669.

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

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

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

Bhat, C.R., A. Sivakumar, and K.W. Axhausen (2003) An analysis of the impact of information and communication technologies on non-maintenance shopping activities, *Transportation Research Part B,* 37 (10) 857-881.

Bierlaire, M. (2002) The network GEV model, presented at the 2002 Swiss Transport Research Conference (STRC), Monte Verita.

Blundell, R. and J. Powell (2001) Endogeneity in semiparametric binary response models, Working Paper, Department of Economics, University College London.

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

Boyd, J. and J. Mellman (1980) The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis, *Transportation Research Part A*, 14 (5-6) 367-378.

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

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

Bresnahan, T.F., S. Stern, 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) S17-S44.

Brownstone, D. (2001) Discrete choice modeling for transportation, in D.A. Hensher (ed.) *Travel Behavior Research: The Leading Edge*, 97-124, Elsevier, Oxford.

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

Brownstone, D. and K. Train (1999) Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, 89 (1-2) 109-129.

Brownstone, D., D.S. Bunch, and K. Train (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles, *Transportation Research Part B*, 34 (5) 315-338.

Brownstone, D., D.S. Bunch, T.F. Golob, and W. Ren (1996) A transactions choice model for forecasting demand for alternative-fuel vehicles, in S. McMullen (ed.) *Research in Transportation Economics*, 4, 87-129, JAI Press Inc.

Brownstone, D., A. Ghosh, T.F. Golob, C. Kazimi, and D. van Amelsfort (2003) Drivers' willingness-to-pay to reduce travel time: Evidence from the San Diego I-15 congestion pricing project, *Transportation Research Part A*, 37 (4) 373-387.

Cardell, S. and F. Dunbar (1980) Measuring the societal impacts of automobile downsizing, *Transportation Research Part A*, 14 (5-6) 423-434.

Chamberlain, G. (1980) Analysis of covariance with qualitative data, *Review of Economic Studies*, 47 (1) 225-238.

Cherchi, E. and J. de D. Ortúzar (2002) Mixed RP/SP models incorporating interaction effects, *Transportation*, 29 (4) 371-395.

Chu, C. (1989) A paired combinational logit model for travel demand analysis, in *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, 4, 295-309, Western Periodicals, Ventura, CA.

Daly, A. and M. Bierlaire (2003) A general and operational representation of GEV models, Technical Report RO-030502, Institute of Mathematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

Daly, A. and S. Zachary (1978) Improved multiple choice models, in D.A. Hensher and M.Q. Dalvi (eds.) *Determinants of Travel Choice*, 335-357, Saxon House, Westmead.

Degeratu, A.M. (1999) Estimation bias in choice models with last choice feedback, *International Journal of Research in Marketing*, 16 (4) 285-306.

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

Goolsbee, A. and A. Petrin (2002) The consumer gains from direct broadcast satellites and the competition with cable tv, Working Paper, Graduate School of Business, University of Chicago.

Greene, W.H., and D.A. Hensher (2003) A latent class model for discrete choice analysis: Contrasts with mixed logit, *Transportation Research Part B*, 37 (8) 681-698.

Gunn, H.F., M.A. Bradley, and D.A. Hensher (1992) High speed rail market projection: Survey design and analysis, *Transportation*, 19 (2) 117-139.

Gupta, S. and P.K. Chintagunta (1994) On using demographic variables to determine segment membership in logit mixture models, *Journal of Marketing Research*, 31 (1) 128-136.

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

Heckman, J.J. (1981) The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, in C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, 179-195, MIT Press, Cambridge.

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

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

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

Hensher, D.A., J.J. Louviere, and J. Swait (1999) Combining sources of preference data, *Journal of Econometrics*, 89 (1-2) 197-221.

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

Hsiao, C. (1986) *Analysis of Panel Data*, Cambridge University Press, Cambridge.

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

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

Jones, P.M. and P.R. Stopher (2003) *Transport Survey Quality and Innovation*, Pergamon, Oxford.

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

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

Kamakura, W.A. and G. Russell (1989) A probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research*, 26 (4) 379-390.

Karlstrom, A. (2001) Developing generalized extreme value models using the Pickands' representation theorem, Working Paper, Infrastructure and Planning, Royal Institute of Technology, Stockholm, Sweden.

Keane, M.P. (1997) Modeling heterogeneity and stated dependence in consumer choice behavior, *American Statistical Association Journal of Business & Economic Statistics*, 15 (3) 310-327.

Koppelman, F. and V. Sethi (2000) Closed-form discrete-choice models, in D.A. Hensher and K.J. Button (eds.) *Handbook of Transport Modelling*, 211-227, Pergamon, New York.

Koppelman, F. and C. Wen (2000) The paired combination logit model: Properties, estimation and application, *Transportation Research Part B*, 34 (2) 75-89.

Koppelman, F.S., C.R. Bhat, and J.L. Schofer (1993) Market research evaluation of actions to reduce traffic congestion: Commuter travel behavior and response to demand reduction actions, *Transportation Research Part A*, 27 (5) 383-393.

Lewbel, A. and O. Linton (2002) Nonparametric censored and truncated regression, *Econometrica*, 70 (2) 765–779.

Louviere, J.J., R.J. Meyer, D.S. Bunch, R. Carson, B. Dellaert, W.M. Hanemann, D.A. Hensher, and J. Irwin (1999) Combining sources of preference data for modeling complex decision processes, *Marketing Letters*, 10 (3) 205-217.

McFadden, D. (1978) Modeling the choice of residential location, in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.) *Spatial Interaction Theory and Planning Models*, 75-96, North-Holland, Amsterdam.

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

Meyer, B.D. (1987) Semiparametric estimation of duration models, Ph.D. Thesis, MIT, Cambridge, Massachusetts.

Meyer, B.D. (1990) Unemployment insurance and unemployment spells, *Econometrica*, 58 (4) 757-782.

Morikawa, T. (1994) Correcting state dependence and serial correlation in the RP/SP combined estimation method, *Transportation*, 21 (2) 153-165.

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

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

Nielsen, O.A., C.O. Hansen, and A. Daly (2000) A large-scale model system for the Copenhagen-Ringsted railway project, presented at the 9th International Association for Travel Behaviour Research Conference, Queensland, Australia.

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

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

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

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

Paag, H., A.J. Daly, and C.L. Rohr (2000) Predicting use of the Copenhagen harbour tunnel, presented at the 9th International Association for Travel Behaviour Research Conference, Queensland, Australia.

Petrin, A. and K. Train (2002) Omitted product attributes in discrete choice models, Working Paper, Department of Economics, University of California, Berkeley.

Revelt, D. and K. Train (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, 80 (4) 647-657.

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

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

Sandor, Z. and K. Train (2004) Quasi-random simulation of discrete choice models, *Transportation Research Part B*, 38 (4) 313-327.

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

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

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

Small, K., C. Winston, and J. Yan (2005) Uncovering the distribution of motorists' preferences for travel time and reliability: Implications for road pricing, *Econometrica*, forthcoming.

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

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

Steinmetz, S. and D. Brownstone (2005) Estimating commuters' "value of time" with noisy data: A multiple imputation approach, *Transportation Research Part B*, in press.

Swait, J. (2001) Choice set generation within the generalized extreme value family of discrete choice models, *Transportation Research B*, 35 (7) 643-666.

Swait, J. and J.J. Louviere (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models, *Journal of Marketing Research*, 30 (3) 305-314.

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

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

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

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

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

Villas-Boas, J.M. and R.S. Winer (1999) Endogeneity in brand choice models, *Management Science*, 45 (10) 1324-1338.

Vovsha, P. (1997) Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area, *Transportation Research Record*, 1607, 6-15.

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

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

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

Williams, H. (1977) On the formation of travel demand models and economic evaluation measures of user benefits, *Environment and Planning A*, 9, 285–344.

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

List of Figure and Table

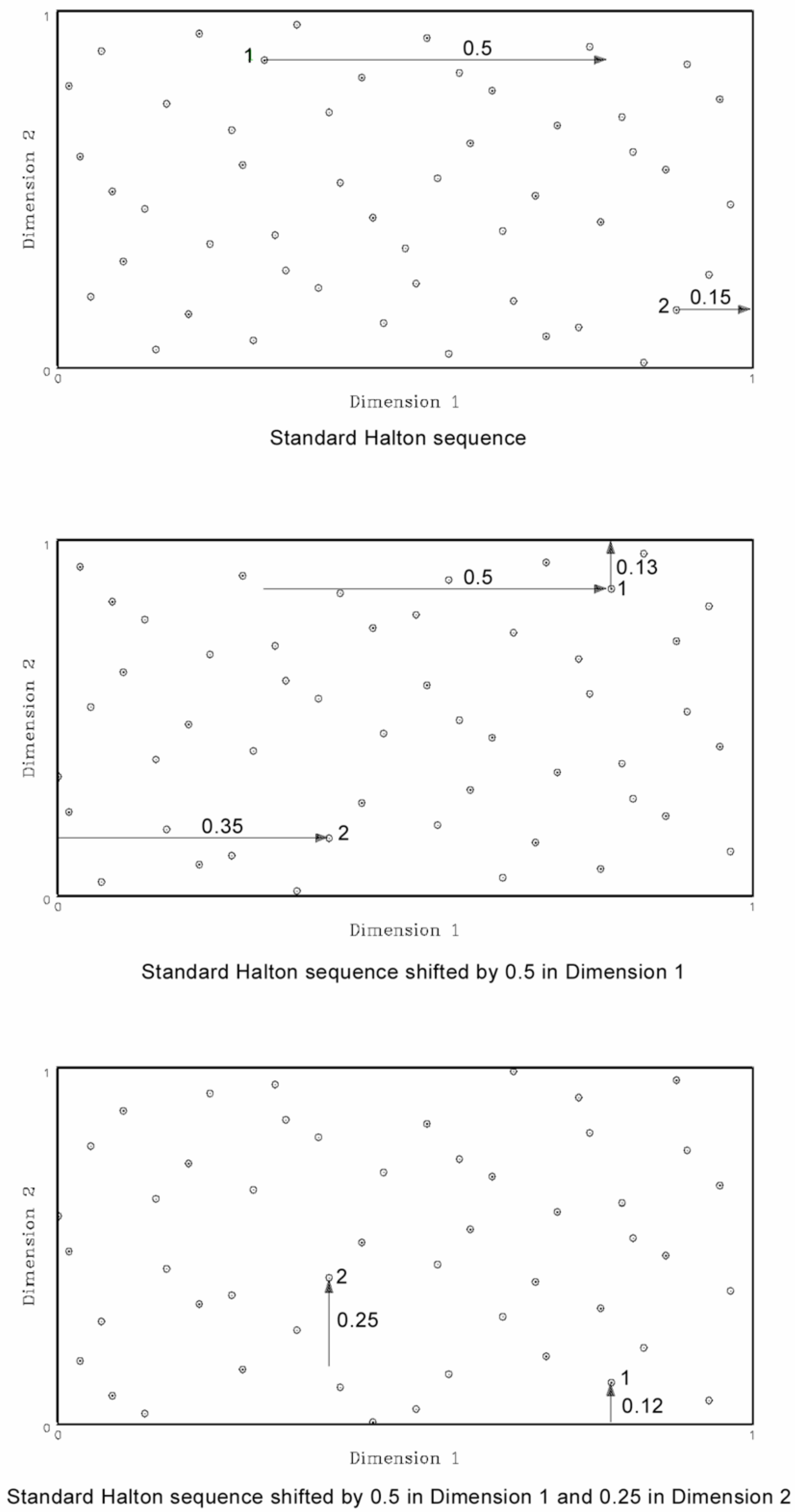

Figure 1. Shifting the Standard Halton Sequence

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

Model Type	Authors	Model Structure	Application Focus	Data Source	Type of Simulation Draws
GEV	Bhat (1998a)	Multinomial Logit-Ordered Generalized Extreme Value (MNL-OGEV)	Travel mode and time-of-day choice for urban shopping trips: Allowing correlations across adjacent time-of-day alternatives and across alternatives sharing the same mode.	1990 RP urban travel survey data collected in the San Francisco metropolitan region.	--
	Koppelman and Wen (2000)	Paired Combinational Logit (PCL)	Intercity travel mode choice: Allowing correlation between each pair of alternatives.	1989 RP intercity travel survey data collected in the Toronto-Montreal corridor.	--
	Swait (2001)	Choice Set Generation Logit (GenL)	Intercity travel mode choice: Accommodating choice set endogeneity through a customized GEV structure.	1987 RP intercity survey collected for travel behavior in Sydney, Canberra, and Melbourne.	--
	Wen and Koppelman (2001)	Generalized Nested Logit (GNL)	Intercity travel mode choice: Allowing general correlation patterns among four modes: car, air, train, and bus.	1989 RP intercity travel survey data collected in the Toronto-Montreal corridor.	--
MMNL	Bhat (1998c)	Error-components structure to accommodate correlation in multiple dimensions (heteroscedasticity across joint choice alternatives is an indirect consequence)	Travel mode and time-of-day choice for social/recreational trips: Allowing unobserved correlation across both time and mode choice.	1990 RP urban travel survey data collected in the San Francisco metropolitan region.	Pseudo-random draws
	Brownstone and Train (1999)	Error components structure to accommodate correlation in a unidimensional context (heteroscedasticity across choice alternatives is an indirect consequence)	Alternative fuel vehicle type choice: Allowing unobserved correlation across non-electric vehicles and based on vehicle size.	1999 SP data collected in California.	Pseudo-random draws
	Brownstone *et al.* (2000)	Error components structure as in Brownstone and Train (1999) and scale differences between revealed and state choice data	Alternative fuel vehicle type choice: Allowing unobserved correlation across non-electric vehicles and based on vehicle size.	1999 SP and RP data collected in California.	Pseudo-random draws
	Bekhor *et al.* (2002)	Error components structure	Travel route choice: Accommodating unobserved correlation on paths with overlapping links.	1997 transportation survey of MIT faculty and staff.	Pseudo-random draws
	Jong *et al.* (2002a)	Error components structure	Travel mode and time-of-day choice: Allowing unobserved correlation across time and mode dimensions.	2001 SP data collected from travelers during extended peak periods (6-11 a.m. and 3-7 p.m.) on weekdays.	Pseudo-random draws

Table 1. continued

MMNL	Bhat (1998b)	Random coefficients structure	Intercity travel mode choice: Accommodating unobserved and observed variations in sensitivity to level-of-service variables.	1989 RP intercity travel survey data collected in the Toronto-Montreal corridor.	Pseudo-random draws
	Bhat (1999)	Random spatial coefficients structure	Urban travel mode choice: Allowing unobserved spatial clustering effects at home-end and work-end due to unobserved zone-specific determinants of mode choice.	1990 RP urban travel survey from the San Francisco Bay area.	Halton draws
	Bhat (2000b)	Random coefficients structure	Urban commute travel mode choice: Accommodating individual-specific unobserved mode preference and level-of-service sensitivities.	1996 RP multiday travel survey data collected in the San Francisco metropolitan region.	Pseudo-random draws
	Han *et al.* (2001)	Random coefficients structure	Travel route choice: Incorporating unobserved individual-specific heterogeneity to route choice determinants (delay, heavy traffic, normal travel time, *etc.*).	2000 SP survey and scenario data collected in Sweden.	Pseudo-random draws
	Hensher (2001)	Random coefficients structure	Long distance travel route choice: Accommodating unobserved individual-specific sensitivities to different components of travel time (free flow time, slowed-down time, and stop time).	2000 SP survey data collected in New Zealand.	Pseudo-random draws
	Brownstone and Small (2005)	Random coefficients structure	Choice of toll versus non-toll facility: Allowing random coefficients to account for individual-specific unobserved preferences, and responsiveness to travel time and unreliability of travel time.	1996-2000 RP/SP survey from the SR-91 facility in Orange County, California.	Pseudo-random draws
	Iragüen and Ortúzar (2004)	Random coefficients structure	Urban route choice: Recognizing unobserved individual heterogeneity in sensitivities to cost, number of accidents, travel time.	2002 SP survey of car users of several private and public employment firms in Santiago.	Information not provided
	Small *et al.* (2005)	Random coefficients structure	Use of toll facilities versus non-toll facilities. Allowing random coefficients to accommodate unobserved individual-specific preferences and sensitivities to cost, travel time, and reliability.	1996-2000 RP/SP survey from the SR-91 facility in Orange County, California.	Pseudo-random draws

Table 1. continued

MMNL	Galilea and Ortúzar (2005)	Random coefficients structure	Residential location choice: Accommodating unobserved individual heterogeneity in sensitivities to travel time to work, monthly rent, and noise level.	2002 SP survey of a sample of Santiago residents.	Information not provided
	Hensher and Greene (2003)	Random coefficients structure	Urban commute travel route choice: Accommodating unobserved individual-specific sensitivities to different components of travel time and cost.	1999 SP survey data sets collected in seven cities in New Zealand.	Halton draws
	Rizzi and Ortúzar (2003)	Random coefficients structure	Urban and interurban route choice: Accommodating unobserved individual heterogeneity in sensitivities to toll, travel time, and accidents.	Stated choice surveys administered to commuters in Santiago in 2002 and to motorists traveling on highways connecting Santiago to Vina del Mar/ Valparaiso/ Rancagua in 1999/2000.	Information not provided
	Silliano and Ortúzar (2005)	Random coefficients structure	Residential choice incorporating unobserved individual heterogeneity in sensitivities to travel time to work, travel time to school, and days of alert status associated with the air quality of the zone of dwelling unit.	2001 SP survey conducted in Santiago.	Information not provided
	Nielsen *et al.* (2000) / Paag *et al.* (2000)	Error components structure, random coefficients	Route choice (toll tunnel or untolled bridge) and time-of-day (peak and offpeak): Accommodating error components across time dimension, random coefficients for sensitivity to time, costs, and intrinsic preference.	SP data collected for the harbor tunnel project in Copenhagen.	Pseudo-random draws
	Jong *et al.* (2002b)	Error components structure, random coefficients	Travel mode and time-of-day choice: Allowing unobserved correlation across time and mode dimensions; individual specific random effects.	2001 SP data collected from travelers during extended peak periods (6-11 a.m. and 3-7 p.m.) on weekdays.	Pseudo-random draws
	Hensher and Greene (2000)	Error components and random coefficients structure	Alternative fuel vehicle type choice: Allowing unobserved correlation across alternative vehicle types through error components and unobserved individual-specific preferences through random effects.	1989 RP/SP data collected in Australia.	Pseudo-random draws

Table 1. continued

MMNL	Bhat and Castelar (2002)	Error components and random coefficients structure	Mode and time-of-day choice: Allowing unobserved correlation across alternatives through error components, preference heterogeneity and variations in responsiveness to level-of-service though random coefficients, and inertia effects of RP choice on SP choices through random coefficients.	1996 RP/SP multiday urban travel survey from the San Francisco Bay area.	Halton draws
	Bhat and Gossen (2004)	Error components and random coefficients structure	Weekend recreational episode type choice: Recognizing unobserved correlation in out-of-home episode type utilities and unobserved individual-specific preferences to participate in in-home, away-from-home, and recreational travel episodes.	2000 RP multiday urban travel survey collected in the San Francisco Bay area.	Halton draws
	Srinivasan and Mahmassani (2003)	Error components and random coefficients structure	Route switching behavior under Advanced Traveler Information System (ATIS): Accommodating error-components associated with a particular decision location in space, unobserved individual-specific heterogeneity in preferences (intrinsic biases) and in age/gender effects.	Simulator-based experiment with Austin area commuters in 2000.	Pseudo-random draws
MGEV and other mixed discrete choice models	Bhat (1999)	Random coefficients superimposed on an ordered response logit structure	Evening commute stop-making behavior: Allowing individual-specific unobserved preferences and responsiveness to commute time.	1990 RP urban travel survey from the San Francisco Bay area.	Halton draws
	Bhat and Zhao (2002)	Random spatial coefficients superimposed on an ordered response base structure	Shopping activity participation: Allowing unobserved residence zone specific effects on number of shopping participations.	1991 RP urban travel survey from the Boston area.	Halton draws
	Bhat and Guo (2004)	Random coefficients with GEV base structure	Residential location choice: Allowing spatial correlation in adjacent spatial units due to unobserved locational factors using a paired Generalized Nested Logit (GNL) structure, and unobserved individual-specific heterogeneity in responsiveness to travel time and other factors.	1996 RP urban travel survey from the Dallas-Fort Worth area.	Halton draws