A Multi-Level Cross-Classified Model for Discrete Response Variables

Chandra R. Bhat

Department of Civil Engineering

University of Texas at Austin

ABSTRACT

In many spatial analysis contexts, the variable of interest is discrete and there is spatial clustering of observations. This paper formulates a model that accommodates clustering along more than one dimension in the context of a discrete response variable. For example, in a travel mode choice context, individuals are clustered by both the home zone in which they live as well as by their work locations. The model formulation takes the form of a mixed logit structure and is estimated by maximum likelihood using a combination of Gaussian quadrature and quasi-Monte Carlo simulation techniques. An application to travel mode choice suggests that ignoring the spatial context in which individuals make mode choice decisions can lead to an inferior data fit as well as provide inconsistent evaluations of transportation policy measures.

Keywords: Mixed logit model, multi-level analysis, spatial analysis, quasi-Monte Carlo sequences, data clustering, Gaussian quadrature, simulation-based econometric estimation, travel mode choice modeling.

1. INTRODUCTION

The recognition and explicit consideration of data clustering has led to several applications of multi-level analysis in the past few years. Data clustering arises in many analysis contexts. For example, individuals are grouped into households, workers are grouped into work places, students are grouped into classes or schools, and residences are grouped into spatial zones. In all these instances, elementary units (individuals, workers, students, and residences) are grouped into higher-level units (households, work places, classes/schools, and spatial zones). Such clustering implies either a natural or acquired differentiation in characteristics of elementary units across higher-level units in addition to variations in elementary units within a higher-level unit. Multi-level analysis recognizes the multiple levels at which the analysis needs to be undertaken; *i.e.*, at the micro-level of elementary units and at the macro-level of higher level units.

Multi-level analysis (also labeled as variance-components analysis or hierarchical modeling) has been applied in several fields, including education (Goldstein *et al*., 1993), sociology (Hox and Kreft, 1994), medicine (Goldstein *et al*., 1994), and geography (Jones and Duncan, 1996). In the past few years, the application of multi-level analysis has been particularly burgeoning in the field of geography since geographical analysis is intrinsically spatial and involves the grouping of elementary units of interest (for example, households and individuals) into higher level spatial clusters (such as neighborhoods, communities, and traffic analysis zones). In such a spatial clustering context, it is important to recognize and preserve between-place heterogeneity (*i.e.*, intrinsic differences across spatial clusters; see Jones and Duncan, 1996). There are several reasons for this. First, spatial autocorrelation is likely to be the norm in geographic analysis since elementary units in the same aggregate spatial unit are likely to be similar in ways not accounted for by observed characteristics of the elementary or aggregate spatial units (Jones and Bullen, 1994). Ignoring such spatial autocorrelation generally results in mis-estimated standard errors in linear models and (in addition) inconsistent parameter estimation in non-linear models. Second, spatial heterogeneity is a natural occurrence in geography with relationships varying in different ways across spatial contexts. Ignoring this heterogeneity can result in structural instability, especially in non-linear models. Third, the spatial units on which data is available for analysis are likely to be a random sample of the population of spatial units. Consequently, the analyst must recognize that the variations in response across spatial units occur because of sampling variance as well as real differences between places. If these two different sources of variation are not distinguished, the result is incorrect estimation of the effect of any explanatory variables defined at the spatial unit (see Bryk and Raudenbush, 1992 for empirical illustrations of this issue). Fourth, heterogeneity among aggregate spatial units and heterogeneity among elementary units needs to be differentiated. As indicated by Jones and Duncan (1996), ignoring this differentiation and modeling the behavior of interest at a single level invites the pitfalls of either the ecological fallacy when the level of analysis is solely at the aggregate spatial level (*i.e.*, failing to recognize that it is the elementary units which act and not aggregate spatial units) or the atomistic fallacy when the analysis is pursued entirely at the elementary unit level (*i.e.*, missing the spatial context in which elementary units behave).

Multi-level analysis satisfies the requirements of a geography that recognizes and accounts for the spatial context within which elementary units act. It accommodates spatial autocorrelation, spatial heterogeneity, higher-level context, and simultaneous handling of the micro-scale of elementary units and the macro-scale of places. Jones and Bullen (1996), Bullen *et al*., 1997, Hox and Kreft (1994), Goldstein (1995), and Jones and Duncan (1995), among other studies, discuss multi-level analysis at length. These studies also discuss the inadequacy of fixed-part expansion methods that attempt to capture spatial variations by including dummy variables for each spatial unit and/or including an observed contextual variable that varies over space. Essentially, fixed-part expansions are equivalent to single-level analysis and ignore the fact that the aggregate spatial units on which data is available for analysis are a random sample of the population of spatial units.

Two important issues associated with multi-level analysis are the functional form of the model structure and the type of clustering. Each of these issues is discussed in the subsequent two sections. Section 1.3 discusses the objective of the paper.

1.1. Functional Form

The functional form of the model depends on the type of dependent variable of interest. When the dependent variable is continuously distributed (such as an achievement score in education or housing price in geography), one can use a model structure that is linear in the fixed and random parts or use a structure that is non-linear in one or both of the fixed and random parts (non-linear model). In the multi-level modeling literature, the linear structure has been the one most commonly used. When the dependent variable is in the form of a proportion (such as a mortality rate or an examination pass rate), the nature and range of the response variable requires the use of a nonlinear relationship between explanatory variables and the response. Typically, a linear predictor of the explanatory variables is developed and a non-linear link function (such as a logit or probit link function) is assumed to relate the expected value of the response variable with the linear predictor. Stochasticity is introduced by assuming a particular distributional form (generally a binomial distribution) for the response variable conditional on the linear predictor. Such a structure is referred to as a generalized linear model since the inverse link function transformation of the dependent variable can be related linearly to the explanatory variables and estimation proceeds in a manner similar to that of a continuous dependent variable (see Goldstein, 1995, Chapter 7 or Neuhaus and Segal, 1997). When the dependent variable is in the form of multiple category proportions or in the form of counts, the generalized linear model structure still applies with minor modifications to the approach with a single proportion response variable.

In the multi-level literature, substantial attention has been centered on the dependent variable types discussed above. However, relatively little (if any) attention has been directed toward the case when a response variable is observed in multiple categories and is intrinsically discrete (rather than being observed as proportions). Examples of such variables of interest include residential location choice of households and work trip travel mode choice of individuals.

1.2. Clustering Type

Multi-level analysis has been, for the most part, restricted to the case when elementary units are nested within one and only one aggregate spatial unit; that is, a strictly hierarchical structure holds. For example, housing units are clustered in a district or students are clustered in schools. However, there are natural instances when elementary units can be classified into more than one higher-level unit. For example, when analyzing the academic performance of students, one might need to consider the clustering based on the school attended as well as by the neighborhood in which the student stays. Such cross-classifications of the elementary units breaks down the hierarchical structure of traditional multi-level analysis and makes model estimation more cumbersome. Goldstein (1987; 1994) discusses random cross-classification models for continuous response variables. But, to the authors' knowledge, equivalent methods for the case of random cross-classification with a discrete response variable have not been formulated and applied.

1.3. Objective of Paper

The objective of this paper is to formulate and apply a methodology that accommodates cross-classification in the context of a multi-level analysis of a multiple-response discrete variable. The model development and application in the paper will be in the context of work travel mode choice. A multi-level situation arises naturally in this context since travel impedances (costs and times) are typically generated at the traffic zone level. Traditional mode choice analysis applies the same zone-to-zone impedance measures to all individuals with the same residence and work zones, thereby not distinguishing between the individual making the trip and the residence/work zones of individuals. Zone-to-zone impedance is considered an attribute of each individual, resulting in a confounding of individual heterogeneity (variations in impedance measures across individuals in the same home/work zones) and place heterogeneity (variations in impedance measures across zonal pairs). It is imperative that we disentangle these two different sources of heterogeneity by formulating a model at the micro-level of individuals as well as the macro-level of traffic zones. In addition, since there are likely to be unobserved factors associated with both an individual's work place as well as residence location that might affect the travel mode decision, we have a classic case of cross-classification (spatial auto-correlation) of individuals by work location as well as by residence zone.

2. THE MODEL STRUCTURE

In the tradition of utility-maximization models, the utility U_{qhwi} that an individual *q* (*q*=1,2,...*Q*) in residence zone *h* (*h*=1,2,...*H*) and work location *w* (*w*=1,2,...*W*) associates with an alternative *i* (*i*=1,2,...*I*) may be written as:

$$U_{qhwi} = \alpha_{hwi} + \beta' x_{qhwi} + \epsilon_{qhwi} \tag{1}$$

where α_{hwi} is a scalar utility term for alternative *i* associated with the residence zone *h* and work location *w* of the individual, x_{qhwi} is a vector of individual-associated variables, β is a corresponding coefficient vector, and ϵ_{qhwi} is an unobserved standard extreme value random term that represents idiosyncratic individual differences in utility after allowing for differences due to observed individual characteristics and zonal-level utility differences. ϵ_{qhwi} is assumed to be independently and identically (IID) distributed (across alternatives and individuals).

Equation (1) represents the micro-level utility model for individuals. We now allow the scalar utility term α_{hwi} to vary across residence and work place zones in a higher-level macro-model:

$$\alpha_{hwi} = \mu' y_i + \gamma' z_{hwi} + \delta_{hi} + \theta_{wi} \tag{2}$$

where y_i is a column vector of 1's and 0's with a 1 in row i and 0 everywhere else, μ is a vector of the "average" effect of unobserved variables on the utilities associated with the modes, z_{hwi} is a vector of observed zonal-level attributes affecting the utility of mode i (for example, travel times and costs), and δ_{hi} and θ_{wi} are random terms that capture unobserved variations across home-zones and work-zones, respectively, in the utility associated with mode *i* (due to identification considerations, y_i will be a column vector of 0's for a base alternative). δ_{hi} and θ_{wi} are assumed to be normally distributed and identically distributed across home zones and work locations, respectively: $\delta_{hi} \sim N(0, \sigma_i^2)$ and $\theta_{wi} \sim N(0, \omega_i^2)$ . .

Next, define $s_{qhwi} = (y_i', z_{hwi}', x_{qhwi}')'$ and $\eta = (\mu', \gamma', \beta')'$. Then, the micro- and macro-models of equations (1) and (2) can be combined to form:

$$U_{qhwi} = \eta' s_{qhwi} + \delta_{hi} + \theta_{wi} + \epsilon_{qhwi} . \tag{3}$$

The usual independence assumptions among all error terms is invoked. Also, δ_{hi} and θ_{wi} are assumed to be independently distributed across home zones and work zones, respectively. Note that if σ_i^2 (variance of δ_{hi}) and ω_i^2 (variance of θ_{wi}) are equal to zero for all modes *i*, then it implies that there are no mode-specific utility differentials across home zones and work places. This situation corresponds to the standard multinomial logit model.

The utility function of equation (3) generates a spatial autocorrelation pattern as follows. For two individuals in the same home zone and work location, the covariance between their utilities for mode *i* is $\sigma_i^2 + \omega_i^2$. For two individuals in the same home zone, but different work locations, the covariance is σ_i^2. For two individuals in the same work location, but different home zones, the covariance is ω_i^2. And for two individuals in different home zones and work locations, the covariance is zero. All cross-mode correlations are, by specification, zero.

In addition to autocorrelation, the utility specification in equation (3) also leads to spatial heteroscedasticity across modes. As discussed in Bhat (1995), heteroscedasticity leads to a competitive structure that does not exhibit the IIA (Independence from irrelevant Alternatives) property of the multinomial logit model. Specifically, as the variance of δ_{hi} and/or θ_{wi} for alternative *i* increases, changes in the systematic component of alternative *i* or of other alternatives are likely to have a smaller impact on the choice share of alternative *i*. Intuitively, we can explain this by noting that the overall error term $\xi_{qhwi}\ (=\delta_{hi}+\theta_{wi}+\varepsilon_{qhwi})$ sets the relative weights of the systematic and uncertain components in estimating the choice probability. The larger the variance of the overall error term of an alternative, the more tempered is the effect (on the choice probability of the alternative) of changes in the systematic utility of that alternative or of other alternatives. This property will be clearly apparent in the policy evaluations conducted in section 4.3.

In utility maximization-based choice modeling, it is only the utility differences that matter in the choice process. Hence appropriate normalization conditions have to imposed on the deterministic part $\eta' s_{qhwi}$ in equation (3). Also, the variance of the home zone and work location random terms must be restricted to zero for a base alternative (equivalently, the random terms δ_{hi} and θ_{wi} must be uniformly zero across home zones and work locations, respectively, for the base alternative). Assume that the zero variance restriction is applied to the first alternative (so, effectively, $\delta_{h1}=0$ for all h and $\theta_{w1}=0$ for all w). Then the other variance terms can be interpreted as a measure of the spatial variation across home zones and work locations in relative utility of other alternatives compared to the first alternative.

Conditional on δ_{hi} and θ_{wi} $(i=2,\ldots,I)$ terms, the probability of choice of mode *i* for individual *q* in residence zone *h* and work location *w* can be written in the usual multinomial logit form (see McFadden, 1978):

$$P_{qhwi}\,|\,(\delta_{h2},\ldots,\delta_{hI},\theta_{h2},\ldots,\theta_{hI}) = \frac{e^{\eta' s_{qhwi}+\delta_{hi}+\theta_{wi}}}{\sum_{j=1}^{I} e^{\eta' s_{qhwj}+\delta_{hj}+\theta_{wj}}}\,. \tag{4}$$

The unconditional probability can be subsequently obtained as:

$$P_{qhwi} = \int_{\delta_{h2}=-\infty}^{+\infty} \cdots \int_{\delta_{hI}=-\infty}^{+\infty} \int_{\theta_{w2}=-\infty}^{+\infty} \cdots \int_{\theta_{wI}=-\infty}^{+\infty} \frac{e^{\eta' s_{qhwi}+\delta_{hi}+\theta_{wi}}}{\sum_{j=1}^{I} e^{\eta' s_{qhwj}+\delta_{hj}+\theta_{wj}}} \, dF(\delta_{h2}) \cdots dF(\delta_{hI}) \, dF(\theta_{w2}) \cdots dF(\theta_{wI}) \,, \tag{5}$$

where F is the univariate cumulative normal distribution.

3. MODEL ESTIMATION

The estimation of the model must recognize the correlation patterns across elementary units generated by the macro-level spatial units of residence zone and work location. The parameters to be estimated include the η vector, the home zone-based standard deviation parameters $\sigma_2, \sigma_3, \ldots, \sigma_I$, and the work location-based standard deviation parameters $\omega_2, \omega_3, \ldots, \omega_I$.

3.1. Likelihood Function Development

The development of the likelihood function for estimation requires additional notation. Define a column vector d_w of dimension W, with each row of the vector being associated with a work location $w' (w' = 1, 2, \ldots, w, \ldots, W)$. . If $w'=w$, the corresponding row of d_w has a value of one; otherwise, the value is zero. Also,define a (Wx1) vector $\theta_i = (\theta_{1i}, \theta_{2i}, \ldots \theta_{Wi})'$. , each of whose elements is distributed $N(0, \omega_i^2)$. In general, and for reasons that will become clear later, the vectors defined above should be associated with the spatial dimension with lower number of spatial units. In the empirical analysis of the current paper, the number of work locations is smaller than the number of residence zones and, therefore, the vectors above are defined with respect to the work location dimension. Equation (4) can now be written as:

$$P_{qhwi} \mid (\delta_{h2}, \ldots \delta_{hI}, \theta_2, \ldots \theta_I) \;=\; \frac{e^{\eta' s_{qhwi} + \delta_{hi} + \theta_i' d_w}}{\sum_{j=1}^{I} e^{\eta' s_{qhwj} + \delta_{hj} + \theta_j' d_w}} \,. \tag{6}$$

Let T_h be the set of individuals q whose residence zone is h. Then, conditional on the scalar δ_{hi} terms (i=2,3,...I) and the vectors $\theta_i\,(i=2,3,...,I)$, the likelihood function for the joint mode choice of all individuals q in home zone h takes the form:

$$\mathcal{L}_h \mid (\delta_{h2}, \ldots, \delta_{hI}, \theta_2, \ldots \theta_I) = \prod_{q \in T_h} \prod_{i} \Big[\;\; P_{qhwi} \mid (\delta_{h2}, \ldots, \delta_{hI}, \theta_2, \ldots \theta_I) \Big]^{M_{qi}}, \tag{7}$$

where M_{qi} is a dummy variable taking the value 1 if individual q chooses mode i and 0 otherwise.

Individuals not in the same residence zone h can have the same work location. So one cannot condition out the θ_i vectors in equation (7) since the equation is specific to a single home zone h. But we can condition out the scalar δ_{hi} terms (i=2,3,...I) to obtain the likelihood function for the joint choice of all individuals q in home zone h conditioned only on the θ_i vectors as:

$$\mathcal{L}_h \mid (\theta_2, \ldots, \theta_I) = \int_{\delta_{h2}=-\infty}^{+\infty} \cdots \int_{\delta_{hI}=-\infty}^{+\infty} \mathcal{L}_h \mid (\delta_{h2}, \ldots, \delta_{hI}, \theta_2, \ldots, \theta_I)\, dF(\delta_{h2}) \ldots dF(\delta_{hI}) \,. \tag{8}$$

Next, the likelihood function for all individuals q across all home zones h (h = 1,2,...H) conditioned on the θ_i vectors takes the following form:

$$\mathcal{L} \mid (\theta_2, \ldots, \theta_I) = \prod_{h=1}^{H} \mathcal{L}_h \mid (\theta_2, \ldots, \theta_I) \,. \tag{9}$$

Finally, the unconditional likelihood function of the entire choice sample can be obtained as:

$$\mathfrak{L} = \int_{\theta_{\mathbf{2}}=-\infty}^{+\infty} \cdots \int_{\theta_{\boldsymbol{I}}=-\infty}^{+\infty} \mathfrak{L} \,|\, (\theta_{\mathbf{2}},\ldots,\theta_{\boldsymbol{I}})\; d\boldsymbol{F}(\theta_{\mathbf{2}})\ldots d\boldsymbol{F}(\theta_{\boldsymbol{I}})$$

$$= \int_{\theta_{\mathbf{2}}=-\infty}^{+\infty} \cdots \int_{\theta_{\boldsymbol{I}}=-\infty}^{+\infty} \prod_{h=1}^{H} \left\{ \int_{\delta_{h2}=-\infty}^{+\infty} \cdots \int_{\delta_{hI}=-\infty}^{+\infty} \prod_{q\in T_h} \prod_{i} \left[\frac{e^{\eta' s_{qhwi}+\delta_{hi}+\theta_{\boldsymbol{i}}' d_w}}{\sum_j e^{\eta' s_{qhwj}+\delta_{hj}+\theta_{\boldsymbol{j}}' d_q}} \right]^{M_{qi}} dF(\delta_{h2})\ldots dF(\delta_{hI}) \right\} d\boldsymbol{F}(\theta_{\mathbf{2}})\ldots d\boldsymbol{F}(\theta_{\boldsymbol{I}}) \;.$$

(10)

where $\boldsymbol{F}$ represents the multivariate normal cumulative distribution.

The inner integration (in parenthesis) in the above equation involves a (I-1)-dimensional integral and is independent of the number of household zones in the sample. The outer integral, on the other hand, involves a (*I*-1)**W*-dimensional integral which is dependent on the number of work locations (note that each $\theta_{\boldsymbol{i}}$ vector is of dimension *W*). Thus, it is most efficient to choose the spatial classification with the higher number of units for the inner integration and the spatial classification with lower number of units for the outer integration.

3.2. Numerical Quadrature and Quasi-Monte Carlo methods

The inner and outer integrals in equation (10) cannot be evaluated analytically because they do not have a closed-form solution. One of two approaches may be adopted for computing the integrals: numerical quadrature methods or Monte-Carlo simulation methods. Numerical quadrature methods generally provide precise approximations at a speed acceptable for maximum likelihood estimation when the dimension of integration is less than 3-4 (see Bratten and Weller, 1979; Sloan and Joe, Chapter 1, 1994). But when the dimension exceeds this, one has to resort to simulation techniques to approximate the integral.

In travel mode choice modeling (and in many other choice contexts), the number of alternatives I is usually less than 4-5. Thus, one can use the quadrature method for the (I-1)-dimensional inner integration. This is particularly helpful because in each function iteration of the likelihood maximization, H inner integrals need to computed. Quadrature is efficient here since use of simulation techniques would require several draws for each of the H integrals as opposed to a single "quadrature draw" corresponding to the multivariate quadrature points.

The outer integrals in equation (10) have to be computed by simulation since its dimensionality will, in general, be greater than 5. Within the framework of simulation methods, one might employ standard Monte Carlo methods (which use pseudo-random sequences) or quasi-Monte Carlo methods (which use quasi-random sequences).

The standard Monte Carlo integration involves the drawing of N points pseudo-randomly in the multi-dimensional domain of integration. It is the most common approach in simulation methods, sometimes refined by variance reduction techniques such as importance sampling or antithetic variates (see Brownstone and Train, 1996, Revelt and Train, 1998 and Bhat, 1998 for recent applications of the standard Monte Carlo integration method). Unfortunately, the convergence rate of the standard Monte Carlo method is rather slow and at a rate of $1/\sqrt{R}$, where R is the number of draws.

The quasi-Monte Carlo method fills the domain of the integration space more uniformly than the standard Monte Carlo method. Basically, quasi-random sequences achieve this through their "clever" sub-random draws of sample points so that the sample points are maximally spread-out (or "maximally" avoiding of each other). The convergence rate for quasi-random sequences is much faster than for the standard Monte Carlo method for a very broad range of integrands. In particular, the upper bound for the asymptotic convergence rate of quasi-random sequences is of the order of $(\ln R)^{d-1}/R$, where d is the dimensionality of the integral (see Niederreiter, 1992; Chapter 4). In the average case, however, Wozniakowski (1991) has shown a much faster convergence rate. Further, Owen (1995, 1997, 1998) has shown that some scrambled versions of quasi-random sequences have a convergence rate of $(\ln R)^{d-1}/R^3$. Computational experiments comparing the performance of standard and quasi-random Monte-Carlo methods have clearly established that the integration error associated with the latter method is significantly less than that associated with the former method for a variety of test functions, variation ranges, and characteristic functions. Thus, for a given error tolerance level, the quasi-random simulation of integrals requires significantly less number of simulation points or "draws" relative to the standard Monte-Carlo method (see Morokoff and Caflisch, 1995; Press *et al.*, 1992, Chapter 7; Brately and Fox, 1988; and Bratley *et al.*, 1992).

The field of quasi-Monte Carlo simulation methods for multi-dimensional integration has been receiving substantial attention in recent years with increasingly sophisticated quasi-random sequences being proposed (see, for example, Sloan and Wozniakowski, 1998; Owen, 1998; Mullen *et al.*, 1995; Kocis and Whiten, 1997). Krommer and Ueberhuber (1994) provide an extensive review of quasi-random sequences. 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). In this paper, we will use the Halton sequence in the quasi-Monte Carlo simulation because of its conceptual simplicity.

3.3. The Halton Sequence

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

$$g = \sum_{l=0}^{L} b_l r^l \text{, where } 0 \le b_l \le r-1 \text{ and } r^L \le g < r^{L+1}. \tag{11}$$

So *g* (*g*=1,2,...*G*) can be represented by the r-*adic* integer string $b_l \ldots b_1 b_0$. The Halton sequence in the prime base r is obtained by taking the radical inverse of g (g=1,2,...G) to the base r by reflecting through the radical point[1]:

$$\varphi_r(g) = 0.b_0 b_1 \ldots b_L \, (base\ r) = \sum_{l=0}^{L} b_l r^{-l-1} \tag{12}$$

[1] A gauss procedure to generate the Halton sequence for a prime base r is available in Feenberg and Skinner (1994).

The sequence above is very uniformly distributed in the interval (0,1) for each prime r. The Halton sequence in S dimensions is obtained by pairing S one-dimensional sequences based on S pairwise relatively prime integers, $r_1, r_2, \ldots, r_S$ (usually the first S primes):

$$\psi_g = (\varphi_{r_1}(g), \varphi_{r_2}(g), \ldots, \varphi_{r_S}(g)) \qquad (13)$$

3.4. Likelihood Evaluation

The evaluation of the likelihood function of equation (10) is discussed in this section. The portion of equation (10) in parenthesis (the inner integration) is the likelihood function for the joint choice of individuals *q* in home zone *h* conditioned on the $\theta_{\boldsymbol{i}}$ vectors (see equations 8 through 10). We can write this inner integration after defining $\varpi_i = \delta_{hi}/(\sqrt{2}\sigma_i)$, $i = 2, 3, \ldots I$, as:

$$\mathfrak{L}_h | (\theta_{\mathbf{2}}, \ldots, \theta_{\boldsymbol{I}}) = \int_{\varpi_2=-\infty}^{+\infty} \cdots \int_{\varpi_I=-\infty}^{+\infty} \prod_{q \in T_h} \prod_i \left[\frac{e^{\eta' s_{qhwi} + \sqrt{2}\sigma_i \varpi_i + \theta_{\boldsymbol{i}}' d_w}}{\sum_j e^{\eta' s_{qhwj} + \sqrt{2}\sigma_j \varpi_j + \theta_{\boldsymbol{j}}' d_q}} \right]^{M_{qi}} e^{-\varpi_2^2} \ldots e^{-\varpi_I^2} \, d\varpi_2 \ldots d\varpi_I. \qquad (14)$$

The above integration is now in an appropriate form for application of a multi-dimensional product formula of the Gauss-Hermite quadrature for given values of $\eta, \sigma_2, \ldots, \sigma_I,$ and the $\theta_{\boldsymbol{i}}$ vectors (see Stroud, 1971).

To evaluate the outer integral in the likelihood function using the Halton sequence, define $\vartheta_{\boldsymbol{i}} = \theta_{\boldsymbol{i}}/\omega_i$, $(i = 2, \ldots, I)$. . Then the likelihood function can be written as:

$$\mathfrak{L} = \int_{\vartheta_{\mathbf{2}}=-\infty}^{+\infty} \cdots \int_{\vartheta_{\boldsymbol{I}}=-\infty}^{+\infty} \prod_{h=1}^{H} \mathfrak{L}_h | (\omega_2 \vartheta_{\mathbf{2}}, \ldots, \omega_I \vartheta_{\boldsymbol{I}}) \, d\Phi(\vartheta_{\mathbf{2}}) \ldots d\Phi(\vartheta_{\boldsymbol{I}}), \qquad (15)$$

where Φ is the standard multivariate cumulative normal distribution. The dimension of integration above is (I-1)*W. The simulation technique to evaluate the integral involves generating the Halton sequence (equation 13) in (I-1)*W dimensions for a specified number of "draws" G (essentially, a Halton "matrix" Y of size G x [(I-1)*W]). However, the Halton sequence is uniformly distributed over the multi-dimensional cube. To obtain the corresponding multivariate normal points over the multi-dimensional domain of the real line, we take the inverse standard normal distribution transformation of Y. By the integral transform result, $X = \Phi^{-1}(Y)$ provides the Halton points for the multi-variate normal distribution (see Fang and Wang, 1994; Chapter 4). The integrand in equation (15) is computed at the resulting points in the columns of the matrix X for each draw (*i.e.*, each row of X) and then the simulated likelihood function is developed in the usual manner as the average of the values of the integrand across the G draws.

4. EMPIRICAL APPLICATION

4.1. Data source and Variable Specification

The data source for the analysis is the San Francisco Bay Area Household Travel Survey conducted by the Metropolitan Transportation Commission (MTC) in the Spring and Fall of 1990 (see White and Company, Inc., 1991 for details of survey sampling and administration procedures).

In this paper, we examine mode choice among three motorized travel modes: drive alone, shared-ride, and transit. The sample comprises 1617 home-based work trip observations associated with the morning home-to-work commute. These trips originate at the home end from 193 different traffic analysis zones with at least five trips originating in each zone. These traffic analysis zones are used as the spatial unit to accommodate the home-end spatial heterogeneity. At the work end, the trips are scattered substantially across 558 traffic analysis zones with most zones attracting one or two trips, and a handful of zones (for example, the San Francisco and San Jose downtown zones) attracting a substantial number of trips. Since the large fraction of zones with one or two trips can cause instability in multi-level modeling, we adopt a typology based on area density (rather than using work zones) to accommodate work-end heterogeneity. The work location types are a) central business districts (CBDs), b) urban business zones, c) urban zones, d) suburban zones, and e) rural zones.

Three level-of-service variables are used in the current analysis: travel cost, travel time and out-of-vehicle travel time over distance. A detailed description of the procedures and assumptions employed in arriving at the level-of-service data is beyond the scope of the current paper, but is available in Purvis (1996). Table 1 presents the mode choice shares and the mean sample statistics for the level-of-service measures.

A number of variables associated with individual socio-demographics and trip characteristics were considered for accommodating observed taste heterogeneity. The final variable specification was determined after systematic testing of several alternative specifications. The variables in the final specification for capturing observed heterogeneity in intrinsic preferences include: a) vehicles per worker in the household, b) household income, and c) an indicator for whether the individual making the commute trip is a non-caucasian or caucasian.

4.2. Empirical Results

In this section, we present the empirical results obtained from applying the multi-level cross-classified MNL (MCMNL) to the mode choice sample. We also estimate a standard multinomial logit (MNL) model and a multi-level hierarchical multinomial logit (MHMNL) model with home-end clustering but no work-end clustering for comparison with the more general MCMNL model.

For the MHMNL model, we found that two quadrature points for each dimension were adequate for approximating the bivariate integration in the likelihood function (we conducted a sensitivity analysis with 2,4, and 10 quadrature points for each dimension and found little difference in the log-likelihood values at convergence and the parameter estimates across the different estimations). For the MCMNL model, we used 2 quadrature points for the inner integration and 100 Halton draws for the simulation estimation of the outer integral (we estimated the log-likelihood function with 25, 50, 75, and 100 Halton draws to examine the sensitivity of results to the number of draws; there was a substantial difference in estimated model parameters between 25 and 50 draws, lesser difference between 50 and 75 draws, and very little difference between 75 and 100 draws).[2]

Table 2 presents the results of the three models. The effects of all explanatory variables are in the same direction in all three models. The coefficients on the socio-demographic variables indicate that individuals from high income earning households are unlikely to use transit. Individuals in households with a high ratio of number of vehicles to number of workers are likely to use the drive alone mode since there is less competition for cars in such households. Non-caucasians appear to be more likely to use the shared-ride mode compared to caucasians.

The level-of-service variables have the expected negative signs. The parameters in the MNL and MHMNL models are about the same, but quite different from those obtained in the MCMNL model. The estimated value of time from the MNL model is \$6.59 per hour for in-vehicle time and \$14.48 per hour for out-of-vehicle time (at the mean one-way distance to work of 9.3 miles). The corresponding values for the MHMNL model (MCMNL model) are \$6.40 (\$7.23) per hour and \$13.68 (\$14.46) per hour, respectively, for in-vehicle and out-of-vehicle times. These results indicate that the implied value of in-vehicle time from the MCMNL model is higher than from the other two models.

Among the spatial heterogeneity parameters, the unobserved variation in utility of the transit mode across home-end zones was not statistically significant and is therefore suppressed in the MHMNL and MCMNL models. In addition, the unobserved variation in utility of the shared-ride mode across work locations was not statistically significant and is therefore suppressed in the MCMNL model. However, the shared-ride utility variance across home zones is statistically significant in the MHMNL and MCMNL models, and the transit utility variance

[2] This result is very encouraging regarding the efficiency of the Halton sequence compared to the random Monte-Carlo approach, where typically 500-1000 or even more draws may be required for accurate simulations. However, a careful and extensive examination of the performance of the Halton sequence *vis-a-vis* the random approach in the context of a multinomial logit kernel is needed to draw more definitive conclusions. The author is currently pursuing such an effort.

across work locations is statistically significant in the MCMNL models. These results indicate that: a) there is no significant difference in the between-zone unobserved heterogeneity for the drive alone utility and the transit utility at the home-end, but the between-zone heterogeneity at the home-end for the shared-ride mode is higher than for the other two modes and b) there is no significant difference in the between-location unobserved heterogeneity for the drive alone utility and the shared-ride utility at the work-end, but the between-location heterogeneity at the work-end for the transit mode is higher than for the other two modes.

The log-likelihood values at convergence for the three models are provided in the last row of the table. Nested likelihood ratio tests among the three models using these log-likelihood values indicate that the MCMNL model provides a superior data fit relative to the other two models. Thus, it appears to be important to accommodate spatial heterogeneity (and the resulting spatial auto-correlation) at both the home and work ends in the current empirical context. Of course, this may not always be the cased in other empirical contexts. However, by structure, the MCMNL model is more general and subsumes the more restrictive models as special cases. Thus, it would be best to estimate the MCMNL model in any empirical context and then settle for a more restrictive structure if the results suggest so.

4.3. Policy Evaluation

In this section, we examine the results of three policy evaluations using the MNL, MHMNL, and MCMNL models. The first policy measure is an increase in drive alone cost by an average of 50 cents due to a congestion pricing measure (the 50 cents average hike corresponds to an increase in drive alone cost of 29.8% for each sample observation). The second is a decrease in transit in-vehicle travel time by an average of 10 minutes (this corresponds to a 38% decrease in transit in-vehicle time for each sample observation). The third is a decrease in transit out-of-vehicle time by an average of 10 minutes (this corresponds to a 33.6% decrease in transit out-of-vehicle time for each sample observation).

The effect of each policy measure on aggregate mode shares is assessed by modifying exogenous variables to reflect a change, computing revised disaggregate probabilities of mode choice using equation (5), calculating revised expected aggregate shares of each mode by sample enumeration, and then obtaining a percentage change from the baseline estimates.

Table 3 provides the results of the policy evaluations. As expected, all models show an increase in non-drive alone mode shares and a decrease in the drive alone mode share for the congestion pricing measure. The models also indicate a decrease in non-transit mode shares and an increase in transit mode share for the transit improvement policies. However, the magnitude of change are very different among the models. For the congestion pricing scheme, the MHMNL model show a substantially lower increase in shared-ride mode share compared to the MNL model. This is an expected result since the MHMNL model accommodates spatial heterogeneity across home-end zones in the shared-ride mode, thus increasing the variance for the shared-ride mode relative to the other two modes (see the discussion in section 3 for the implication of differential utility variances across alternatives on competitive structure). The MCMNL model provides dramatically different results from the other two models. Compared to the MNL model, the MCMNL model indicates a substantially lower draw away from the drive alone mode and to the non-drive alone modes. This is, of course, because of the home-end spatial heterogeneity in the transit mode utility combined with the work-end spatial heterogeneity in the transit mode utility. Compared to the MHMNL model, the MCMNL model allows a much lower shift to the transit mode because the latter heterogeneity component is ignored by the MHMNL model. For the transit improvement policies, we see a similar pattern of a small draw from the drive alone mode to the transit mode in the MCMNL model compared to the other two models.

To summarize, the substantive implications for policy analysis from the MNL, MHMNL, and MCMNL models are quite different in the current empirical context. Specifically, the MNL and MHMNL models overestimate the potential traffic congestion alleviation due to auto-use dis-incentives (such as congestion pricing) and non-drive alone use incentives (such as improvements in transit service). Thus, ignoring the spatial context (or clustering) of individuals into home-end zones and work-end locations has the potential to lead to misleading evaluations and misinformed policy implementations.[3]

5. SUMMARY AND CONCLUSIONS

This paper formulates and applies a model that accommodates cross-classification in the context of a multi-level analysis of a discrete response variable. The model takes the form of a mixed logit structure that incorporates spatial heterogeneity at both the home-end and the work-end within the context of urban work travel mode choice.

The likelihood function for model estimation includes the computation of an (I-1) dimensional integral (I being the number of choice alternatives) nested within a (I-1)*W-dimensional integral (W being the number of work-end location types). The former integral is computed using a Gaussian quadrature technique because of its relatively low dimensionality, while the latter integral is evaluated using a simulation approach because of its high dimensionality. Unlike previous applications of a (pseudo-) random Monte Carlo procedure for simulation in the mixed logit literature, the current paper uses a quasi-random Monte Carlo procedure based on the Halton sequence. The results suggest that about 100 draws from the Halton sequence are sufficient for the multi-dimensional integration in the current empirical context. This is very encouraging and suggests more widespread consideration of quasi-random sequences in simulation-based econometric methods. It is indeed quite surprising that while there is a substantial body of theoretical and computational literature in physics and mathematics extolling the much faster convergence rate and superior accuracy of quasi-Monte Carlo methods over traditional (pseudo-) random Monte Carlo methods, the quasi-methods have seldom been applied in econometric literature (to the author's knowledge, the only other study

[3] These statements assume that the MCMNL model is more reflective of reality than the other models; the assumption is based on the superior data fit of the MCMNL model compared to the MNL and MHMNL models.

in econometrics that uses quasi-methods is the time series analysis by Feenberg and Skinner, 1994).

The application of the cross-classified multi-level model to travel mode choice suggests that it is important to accommodate the spatial context in which individuals make such decisions. Failure to recognize the spatial context can lead to a diminished data fit as well as misleading evaluations of traffic control measures aimed at alleviating traffic congestion.

ACKNOWLEDGEMENTS

The author benefitted considerably from discussions and e-mail exchanges with Jonathan Skinner on the application of the Halton sequence. The research was supported by National Science Foundation grants DMS 9208758 and DMS 9313013 to the National Institute of Statistical Sciences (NISS). The constructive suggestions of two reviewers on an earlier draft of the paper are greatly appreciated.

REFERENCES

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

Bhat, C.R. (1998) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling, *Transportation Research*, 32B, 21-43.

Bratley, P. and B.L. Fox (1988) Implementing Sobol's quasi-random sequence generator, *ACM Transactions on Mathematical Software*, 14, 88-100.

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

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

Brownstone, D. and K. Train (1996) Forecasting new product penetration with flexible substitution patterns, working paper, Department of Economics, University of California, Irvine.

Bryk, A.S. and S.W. Raudenbush (1992) *Hierarchical Linear Models: Application and Data Analysis methods*, Sage Publications, Beverly Hills, CA.

Bullen, N., K. Jones and C. Duncan (1997) Modeling complexity: analyzing between-individual and between-place variation - a multilevel tutorial, *Environment and Planning*, 29A, 585-609.

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

Feenberg, D. and J. Skinner (1994) The risk and duration of catastrophic health expenditure, *The review of Economics and Statistics*, LXXVI, 633-647.

Goldstein, H., M.J.R. Healy and J. Rasbash (1994) Multilevel time series models with application to repeated measures data, *Statistics in Medicine*, 13, 1643-55.

Goldstein, H., J. Rasbash, M. Yang, G. Woodhouse, H. Pan, D. Nuttall and S. Thomas (1993) A multilevel analysis of school examination results, *Oxford Review of education*, 19, 425-433.

Goldstein, H. (1995) *Multilevel Statistical Models*, Second Edition, Edward Arnold, London.

Goldstein, H. (1987) Multilevel covariance component models, *Biometrika*, 74, 430-431.

Goldstein, H. (1994) Multilevel cross-classification models, *Sociological Methods and Research*, 22, 364-375.

Hox, J.J. and I.G. Kreft (1994) *Multilevel analysis methods, Sociological Methods and Research*, 22, 283-299.

Jones, K. and C. Duncan (1995) Individuals and their ecologies: analyzing the geography of chronic illness within a multilevel modeling framework, *Health and Place*, 1, 27-40.

Jones, K. and C. Duncan (1996) People and places: the multilevel model as a general framework for the quantitative analysis of geographical data, in *Spatial Analysis: Modelling in a GIS Environment*, editors: P. Longley and M. Batty, 79-104, GeoInformation International, Cambridge.

Jones, K. and N. Bullen (1994) Contextual models of urban home prices: a comparison of fixed and random coefficient models developed by expansion, *Economic Geography*, 70, 252-272.

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

Krommer, A.R. and C.W. Ueberhuber (1994) *Numerical integration on advanced computer systems*, Springer-Verlag, Berlin, Germany.

Morokoff, W.J. and R.E. Caflisch (1995) Quasi-Monte Carlo integration, *Journal of Computational Physics*, 122, 218-230.

Mullen, G.L., A. Mahalanabis, and H. Niederreiter (1995) Tables of (T,M,S)-net and (T,S)-sequence parameters, in *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, editors: H. Niederreiter and J-S. Shiue, 58-86, Springer, New York.

Nuehaus, J.M. and M.R. Segal (1997) An assessment of approximate maximum likelihood estimators in generalized linear models, in *Modelling Longitudinal and Spatially Correlated Data*, editors: T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren, R.D. Wolfinger, 11-22, Springer, New York.

Niederreiter, H. (1992) *Random number generation and quasi-Monte Carlo methods*, 63, CBMS-NSF Regional Conference Series in Applied Math., SIAM, Philadelphia, PA.

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

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

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

Press, W.H., S.A. Teukolsky and M. Nerlove (1992) Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Massachusetts

Purvis, C. (1996) Estimation of home-based work mode choice models, Technical Memorandum in San Francisco Bay Area 1990 Travel Demand Model Development Project, Planning Section, Bay Area Metropolitan Transportation Commission.

Revelt, D. and K. Train (1997) Mixed logit with repeated choices: households' choices of appliance efficiency level, *forthcoming*, *Review of Economics and Statistics*.

Sloan, I.H. and S. Jow (1994) Lattice methods for multiple integration, Clarendon Press, Oxford University Press

Sloan, J.H. and H. Wozniakowski (1998) When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *Journal of Complexity*, 14, 1-33.

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

White, E.H. and Company, Inc. (1991) 1990 bay area travel survey: final report, submitted to the Metropolitan Transportation Commission, Oakland, California.

Wozniakowski, H. (1991) Average case complexity of multivariate integration, *Bulletin of the American Mathematical Society*, 24, 185-194.

Table 1: Mode Choice Shares and Mean Sample Values of Level-of-Service Attributes

Mode	Choice shares	In-vehicle travel time in mins.	Out-of-vehicle travel time in mins.	Cost in cents
Drive alone	0.756	17.3	4.1	167.6
Shared ride	0.141	22.2	4.1	83.8
Transit	0.103	26.3	29.7	129.8

Table 2: Urban Work Mode Choice Estimation Results

Variable	Multinomial Logit Model (MNL)		Multi-Level Hierarchical MNL (MHMNL)		Multi-Level Cross-Classified MNL (MCMNL)	
	Parameter	t-stat.	Parameter	t-stat.	Parameter	t-stat.
Alternative specific constants (Drive alone or DA mode is base)						
Shared-ride (SR)	-0.918	-4.87	-1.206	-5.21	-1.147	-4.96
Transit (TR)	1.348	4.07	1.345	4.23	0.456	1.16
Socio-demographic variables						
Income in 000's (specific to TR)	-0.009	-2.58	-0.009	-2.49	-0.009	-2.81
Vehicles per worker (specific to DA)	0.859	6.33	0.871	6.84	0.860	6.18
Non-caucasian (specific to SR)	0.703	4.61	0.643	4.07	0.671	3.96
Level-of-service variables						
Total travel time (minutes)	-0.047	-7.25	-0.048	-8.59	-0.033	-4.42
Out-of-vehicle time over distance (mins./mile)	-0.313	-4.47	-0.302	-6.96	-0.180	-3.12
Travel cost (dollars)	-0.430	-9.98	-0.451	-10.53	-0.270	-5.14
Spatial heterogeneity parameters						
SR utility variance across home zones	-		0.945	6.63	0.932	6.51
TR utility variance across work locations	-		-		2.040	4.83
Log-likelihood at convergence	-900.31		-887.22		-879.35	

Note: The log-likelihood at equal shares is -1776.46 and at sample shares is -1168.06. The number of observations in the sample is 1617.

Table 3: Results of Policy Evaluation

Policy scenario	Travel mode	Percentage change in aggregate share due to policy action from...		
		MNL Model	MHMNL Model	MCMNL Model
50 cents increase (on average across the sample) in drive alone cost	Drive alone	-4.1	-3.4	-0.9
	Shared-ride	+12.3	+1.1	+0.6
	Transit	+13.4	+17.03	+0.4
10 minutes decrease (on average across the sample) in transit in-vehicle time	Drive alone	-5.2	-5.7	-2.4
	Shared-ride	-8.5	-0.8	-1.1
	Transit	+49.3	+44.4	+5.3
10 minutes decrease (on average across the sample) in transit out-of-vehicle time	Drive alone	-3.3	-4.0	-1.0
	Shared-ride	-7.1	-0.7	-0.5
	Transit	+33.5	+31.7	+2.2

List of Tables