A mixed spatially correlated logit model: formulation and application to residential choice modeling

Chandra R. Bhat *, Jessica Guo

Department of Civil Engineering, The University of Texas at Austin, ECJ Hall, Suite 6.8, Austin, TX 78712, USA

Received 12 August 2002; received in revised form 12 December 2002; accepted 13 December 2002

Abstract

In recent years, there have been important developments in the simulation analysis of the mixed multinomial logit model as well as in the formulation of increasingly flexible closed-form models belonging to the generalized extreme value class. In this paper, we bring these developments together to propose a mixed spatially correlated logit (MSCL) model for location-related choices. The MSCL model represents a powerful approach to capture both random taste variations as well as spatial correlation in location choice analysis. The MSCL model is applied to an analysis of residential location choice using data drawn from the 1996 Dallas–Fort Worth household survey. The empirical results underscore the need to capture unobserved taste variations and spatial correlation, both for improved data fit and the realistic assessment of the effect of sociodemographic, transportation system, and land-use changes on residential location choice.
© 2003 Elsevier Ltd. All rights reserved.

1. Introduction

Discrete choice models have a long history of application in the economic, transportation, marketing, and geography fields, among other areas. Most discrete choice models are based on the random utility maximization (RUM) hypothesis. Within the class of RUM-based models, the multinomial logit (MNL) model has been the most widely used structure. The random components of the utilities of the different alternatives in the MNL model are assumed to be independent and identically distributed (IID) with a type I extreme value (or Gumbel) distribution (Johnson and Kotz, 1970; Chapter 21). In addition, the responsiveness to attributes of alternatives across individuals is assumed to be homogeneous after controlling for observed individual characteristics

---

* Corresponding author. Tel.: +1-512-471-4535; fax: +1-512-475-8744.
E-mail addresses: bhat@mail.utexas.edu (C.R. Bhat), jessica.guo@mail.utexas.edu (J. Guo).

(i.e., the MNL model maintains an assumption of unobserved response homogeneity). For example, in a mode choice model, the MNL model maintains the same utility parameters on the level-of-service (LOS) attributes across observationally identical individuals. These foregoing two assumptions together lead to the simple and elegant closed-form mathematical structure of the MNL. However, the assumptions also leave the MNL model saddled with the "independence of irrelevant alternatives" (IIA) property at the individual level (Luce and Suppes, 1965; Ben-Akiva and Lerman, 1985).

There are several ways to relax the IID error structure and/or the unobserved response homogeneity assumption. The IID error structure assumption can be relaxed in one of three ways: (a) allowing the random components to be correlated while maintaining the assumption that they are identically distributed (identical, but non-independent, random components), (b) allowing the random components to be non-identically distributed, but maintaining the independence assumption (non-identical, but independent, random components), or (c) allowing the random components to be non-identical and non-independent (non-identical, non-independent, random components). Unobserved response homogeneity may be relaxed in one of two ways: (a) allowing the attribute coefficients to vary randomly due to unobserved factors using a continuous distribution across individuals (random-coefficients approach), or (b) allowing the attribute coefficients to vary randomly due to unobserved factors using a non-parametric discrete distribution across individuals (latent segmentation approach). Within each of the different approaches to relax the IID and unobserved response homogeneity assumptions, there are several different types of model structures that may be used (see Bhat, 2002a for a detailed discussion). Of these model structures, two classes of models have received particular attention, corresponding to the generalized extreme value (GEV) class of models and the mixed multinomial logit (MMNL) class of models. These two classes of models are discussed in turn in Sections 1.1 and 1.2. Section 1.3 discusses the motivation for combining the GEV class of models with the MMNL class of models in certain empirical circumstances.

1.1. The GEV class of models

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

The general structure of the GEV class of models was derived by McFadden (1978) from the random utility maximization hypothesis, and generalized by Ben-Akiva and Francois (1983). Several specific GEV models have been formulated and applied, 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, 1998), and the product differentiation logit (PDL) model (Bresnahan 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 labels the choice set generation ogit (GenL) model. Wen and Koppelman's derivation of the GNL model is motivated from the perspective of flexible substitution patterns across alternatives, while Swait's derivation of the GenL model is motivated from the concept of latent choice sets of individuals. Wen and Koppelman (2001) illustrate the general nature of the GNL formulation by deriving the other GEV models 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. Bierlaire (2002) has built on this concept and has proposed a very general network structure-based motivation and design of GEV models, which he refers to as the network GEV model.

The GNL model proposed by Wen and Koppelman (2001) is conceptually appealing from a formulation standpoint and allows substantial flexibility. However, in practice, the flexibility of the GNL model can be realized only if one is able and willing to estimate a large number of dissimilarity and allocation parameters. The net result is that the analyst will have to impose informed restrictions on the general GNL model formulation that are customized to the application context under investigation.

The advantage of all the GEV models discussed above is that they allow relaxations of the independence assumption among alternative error terms, while maintaining closed-form expressions for the choice probabilities.

1.2. The MMNL class of models

The MMNL class of models is a generalization of the MNL model. It involves the integration of the MNL formula over the distribution of unobserved random parameters. It takes the structure shown below:

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

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

The MMNL model structure of Eq. (1) can be motivated from two very different (but formally equivalent) perspectives (see Bhat, 2000). 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). Most importantly, the MMNL class of models can approximate any discrete choice model derived from RUM (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 Green, 2001). 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, 2002a).

1.3. 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 (MGEV) class may appear unnecessary. However, there are instances when substantial computational efficiency gains may be achieved using a MGEV structure. Consider, for instance, a model for household residential location choice. It is possible, if not very likely, that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow alternatives that are contiguous to be correlated. In the MMNL structure, such a correlation structure will require the specification of as many error components as the number of pairs of spatially contiguous alternatives. [1] On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation. However, the GEV model structure cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such heterogeneity, leading to a parsimonious and powerful MGEV structure.

This paper proposes a mixed spatially correlated logit (MSCL) model that uses a GEV-based structure to accommodate correlation in the utility of spatial units, and superimposes a mixing distribution over the GEV structure to capture unobserved response heterogeneity. The GEV structure used in the paper 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). The MSCL model developed in this paper emphasizes the fact that closed-form GEV-based models and open-form mixed distribution models are not as mutually exclusive as may be the impression in the discrete choice field.

The rest of the paper is structured as follows. The next section discusses the structure, properties, and estimation of the MSCL model. Section 3 discusses an empirical application of the MSCL model to residential location choice. The final section summarizes the important findings from the study.

2. Model structure and properties

In this section, we first maintain the assumption of observed response homogeneity and propose the spatially correlated logit (SCL) model (Sections 2.1 and 2.2). Subsequently, we relax the

---

[1] In fact, the multinomial probit model (MNP) is a more efficient formulation to capture spatial correlation (see Bolduc, 1992; Garrido and Mahmassani, 2000 for examples of the applications of the MNP model to accommodate spatial correlation). However, even the MNP model entails multidimensional integration of the order of the number of spatial units in the analysis.

assumption of unobserved response homogeneity to develop the MSCL model and present the estimation procedure for the MSCL model (Sections 2.3 and 2.4).

2.1. Notation and definitions

Consider a household residential choice decision among I spatial units ($i = 1, 2, \ldots, I$). Let ω_{ij} be a dummy variable that takes a value of 1 if zone j is adjacent to zone i and 0 otherwise (by convention, $\omega_{ii} = 0$). As indicated in the previous section, a common specification in the spatial analysis literature for capturing spatial correlation is to allow immediately contiguous alternatives to share common unobserved elements. Let ρ be a dissimilarity parameter capturing the correlation between contiguous spatial units.

With the above definitions and notations, the number of spatial units adjacent to spatial unit i is $\sum_{j=1}^{I} \omega_{ij}$; that is, spatial unit i has an unobserved shared component with $\sum_{j=1}^{I} \omega_{ij}$ other spatial units. This unobserved correlation may be represented in the form of paired nests with dissimilarity parameter ρ, each nest including the alternative i and one of its adjacent spatial units. The total number of paired nests is $\sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \omega_{ij}$.

We next define an allocation parameter $\alpha_{i,ij}$ representing the allocation of alternative i to the paired nest with alternatives i and j. Intuitively, the larger is the allocation of alternative i to nest ij, the greater is the correlation generated between alternatives i and j. Since there is no reason to believe that an alternative i is going to be more correlated with any one neighboring unit compared to other neighboring units, we maintain the assumption of equal allocation of alternative i to each paired nest comprising i and one of its adjacent spatial units. Further, since the sensitivity to changes in neighboring spatial units can be expected to be larger if an alternative i is contiguous to fewer spatial units than if it is contiguous to several spatial units, we define the allocation parameter for alternative i as:

$$\alpha_{i,ij} = \frac{\omega_{ij}}{\sum_k \omega_{ik}}. \tag{2}$$

The above definition satisfies the condition $\sum_j \alpha_{i,ij} = 1$. That is, the total allocation of alternative i across all pairings of i with other alternatives (both contiguous and non-contiguous) is unity.

We will now consider a simple case of five spatial units to clarify the notations and the analysis setup. Consider the configuration shown toward the top of Fig. 1. To generate spatial correlation, we define seven spatial nests, as indicated in the middle diagram of Fig. 1. The corresponding contiguity matrix and allocation parameters are provided in the table toward the bottom of Fig. 1.
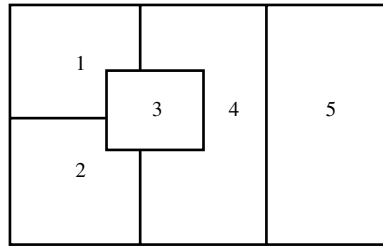
2.2. The SCL model

Consider the following G function within the class of GEV models:

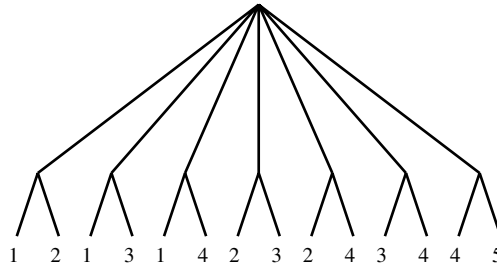$$G(y_1, y_2, \ldots, y_I) = \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} [(\alpha_{i,ij} y_i)^{1/\rho} + (\alpha_{j,ij} y_j)^{1/\rho}]^{\rho}, \tag{3}$$

where $0 < \alpha_{i,ij} < 1$ for all i and j, $0 < \rho \leqslant 1$, $y_i > 0$, for all i, and $\sum_j \alpha_{i,ij} = 1$ for all i. Then it is easy to verify that G is non-negative, homogeneous of degree one, tending toward $+\infty$ when any

(a) Spatial configuration

(b) Paired nested structure for generating spatial correlation between adjacent spatial units

(c) Contiguity (Allocation) Matrix

Alternative	Spatial Units										Number of Adjacent Spatial Units
	1		2		3		4		5		
1	0	(0)	1	(0.33)	1	(0.33)	1	(0.33)	0		3
2	1	(0.33)	0	(0)	1	(0.33)	1	(0.33)	0		3
3	1	(0.33)	1	(0.33)	0	(0)	1	(0.33)	0		3
4	1	(0.25)	1	(0.25)	1	(0.25)	0	(0)	1	(0.25)	4
5	0		0		0		1	(1)	0	(0)	1

Fig. 1. A simple example of residential choice among five spatial units.

argument tends toward $+\infty$, and whose nth cross-partial derivatives are non-negative for odd n and non-positive for even n because $0 < \rho < 1$. Thus, the following function represents a cumulative extreme-value distribution

$$F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_I) = \exp\left\{ -\sum_{i=1}^{I-1} \sum_{j=i+1}^{I} [(\alpha_{i,ij}e^{-\varepsilon_i})^{1/\rho} + (\alpha_{j,ij}e^{-\varepsilon_j})^{1/\rho}]^{\rho} \right\}. \tag{4}$$

Let ε_i in the above equation represent the random element of utility for spatial unit i. The marginal cumulative distribution function (CDF) of each stochastic element ε_i is univariate extreme-value as follows:

$$F(\varepsilon_i) = \exp\left\{ -\sum_{j \neq i} \alpha_{i,ij}e^{-\varepsilon_i} \right\} = \exp\{-e^{-\varepsilon_i}\}, \tag{5}$$

which is the standard Gumbel distribution function. The bivariate marginal CDF for two stochastic elements ε_j and ε_k of two non-adjacent spatial units i and k is as follows:

$$H(\varepsilon_i, \varepsilon_k) = \exp\left\{ -\sum_{j\neq i}\alpha_{i,ij}\mathrm{e}^{-\varepsilon_i} - \sum_{j\neq k}\alpha_{k,ij}\mathrm{e}^{-\varepsilon_k} \right\} = \exp\{-\mathrm{e}^{-\varepsilon_i} - \mathrm{e}^{-\varepsilon_k}\}, \tag{6}$$

which represents the case of independence of the two non-adjacent spatial units. Finally, the bivariate marginal CDF for two stochastic elements ε_i and ε_k of two adjacent spatial units i and k is as follows:

$$H(\varepsilon_i, \varepsilon_k) = \exp\{-(1 - \alpha_{i,ik})\mathrm{e}^{-\varepsilon_i} - (1 - \alpha_{k,ik})\mathrm{e}^{-\varepsilon_k} - [(\alpha_{i,ik}\mathrm{e}^{-\varepsilon_i})^{1/\rho} + (\alpha_{k,ik}\mathrm{e}^{-\varepsilon_k})^{1/\rho}]^{\rho}\}. \tag{7}$$

When the dissimilarity parameter ρ is equal to 1 (that is, when there is no spatial correlation between any adjacent spatial units), the above function collapses to Eq. (6), which is the case of independence among spatial alternatives (or the MNL model). In the general case when $0 < \rho \leqslant 1$, the correlation between two spatial units cannot be written in closed form. This correlation can be computed using numerical integration by noting that the marginal bivariate probability density function associated with the CDF in Eq. (7) is:

$$f(\varepsilon_i, \varepsilon_k) = H(\varepsilon_i, \varepsilon_k)[B_{ik}C_{ik} + D_{ik}], \text{ where}$$

$$A_{ik} = (\alpha_{i,ik}\mathrm{e}^{-\varepsilon_i})^{1/\rho} + (\alpha_{k,ik}\mathrm{e}^{-\varepsilon_k})^{1/\rho},$$

$$B_{ik} = [(1 - \alpha_{i,ik})\mathrm{e}^{-\varepsilon_i} + A_{ik}^{\rho-1}(\alpha_{i,ik}\mathrm{e}^{\varepsilon_i})^{1/\rho}],$$

$$C_{ik} = [(1 - \alpha_{k,ik})\mathrm{e}^{-\varepsilon_k} + A_{ik}^{\rho-1}(\alpha_{k,ik}\mathrm{e}^{-\alpha_k})^{1/\rho}], \tag{8}$$

$$D_{ik} = \left(\frac{1-\rho}{\rho}\right)A_{ik}^{\rho-2}(\alpha_{k,ik}\mathrm{e}^{-\varepsilon_k})^{1/\rho}(\alpha_{i,ik}\mathrm{e}^{-\varepsilon_i})^{1/\rho}.$$

Table 1 presents the correlations between the stochastic elements of two adjacent spatial units i and j for different values of the dissimilarity parameter ρ and for different numbers of spatial units adjacent to spatial units i and j. Due to the symmetric nature of the matrix, only the upper triangle is presented in the table. Two important aspects of this correlation matrix may be readily observed. First, the more the number of spatial units that j is adjacent to, the less is the correlation between i and j. Second, if spatial unit i is adjacent to several alternatives, the impact of a change in i is spread out across its many adjacent alternatives. This leads to a smaller correlation between i and any of its adjacent alternatives j. Third, for any given number of spatial units adjacent to i and j, the correlation between i and j decreases as the dissimilarity parameter increases. As expected, the correlation between adjacent pairs of alternatives is zero when $\rho = 1$. This corresponds to the MNL model.

To obtain the probability of choice for each spatial alternative i in the SCL model, consider a utility maximum decision process where the utility of each spatial unit (U_i) is written in the usual form as the sum of a deterministic component (V_i) and a random component ε_i. If the random components follow the CDF in Eq. (4), then, by the GEV postulate, the probability of choosing the jth spatial unit is:

Table 1
Correlation between stochastic utilities of adjacent spatial units i and j

Dissimilarity parameter ρ	Number of spatial units adjacent to alternative i (including j)	Number of spatial units adjacent to alternative j (including i)				
		1	2	3	4	5
0.2	1	0.97	0.62	0.48	0.39	0.34
	2		0.46	0.34	0.30	0.26
	3			0.30	0.24	0.21
	4				0.22	0.19
	5					0.18
0.4	1	0.84	0.55	0.43	0.35	0.30
	2		0.39	0.31	0.27	0.23
	3			0.26	0.22	0.19
	4				0.19	0.17
	5					0.15
0.6	1	0.64	0.43	0.34	0.28	0.24
	2		0.30	0.25	0.21	0.18
	3			0.20	0.17	0.15
	4				0.15	0.13
	5					0.12
0.8	1	0.36	0.25	0.19	0.16	0.14
	2		0.17	0.14	0.12	0.11
	3			0.12	0.10	0.09
	4				0.09	0.08
	5					0.07
1.0	1	0.00	0.00	0.00	0.00	0.00
	2		0.00	0.00	0.00	0.00
	3			0.00	0.00	0.00
	4				0.00	0.00
	5					0.00

$$
\begin{aligned}
P_i &= \frac{\sum_{j \neq i} (\alpha_{i,ij} e^{V_i})^{1/\rho} [(\alpha_{i,ij} e^{V_i})^{1/\rho} + (\alpha_{j,ij} e^{V_j})^{1/\rho}]^{\rho-1}}{\sum_{k=1}^{I-1} \sum_{l=i+1}^{I} [(\alpha_{k,kl} e^{V_k})^{1/\rho} + (\alpha_{l,kl} e^{V_l})^{1/\rho}]^{\rho}} \\
&= \sum_{j \neq i} \frac{(\alpha_{i,ij} e^{V_i})^{1/\rho}}{(\alpha_{i,ij} e^{V_i})^{1/\rho} + (\alpha_{j,ij} e^{V_j})^{1/\rho}} \times \frac{[(\alpha_{i,ij} e^{V_i})^{1/\rho} + (\alpha_{j,ij} e^{V_j})^{1/\rho}]^{\rho}}{\sum_{k=1}^{I-1} \sum_{l=i+1}^{I} [(\alpha_{k,kl} e^{V_k})^{1/\rho} + (\alpha_{l,kl} e^{V_j})^{1/\rho}]^{\rho}} \\
&= \sum_{j \neq i} P_{i|ij} \times P_{ij}.
\end{aligned}
\tag{9}
$$

The self- and cross-elasticities for the MNL model and the SCL model are provided in the first two rows of Table 2 (the elasticity expressions assume a linear-in-parameters form for V_i; that is, $V_i = \beta' x_i$). The cross-elasticity for the MNL model reflects the IIA property (equal cross-elasticities of the effect of alternative i on any alternative j). The cross-elasticity expression in the SCL model indicates equal proportionate change in all non-adjacent alternatives to i due to a change in the utility of alternative i. However, the cross-elasticities are higher for spatial units contiguous to i.

Table 2
Expressions for the direct and cross-elasticities in the MNL, SCL and MSCL models

Model	Direct elasticity[a]	Cross-elasticity[b]
MNL	$(1 - P_i)\beta_m x_{im}$	$-P_i\beta_m x_{im}$
SCL	$\left\{ \sum_{j \neq i} P_{i|ij} P_{ij} \left[(1 - P_i) + \left(\frac{1}{\rho} - 1 \right)(1 - P_{i|ij}) \right] \right\} \frac{\beta_m x_{im}}{P_i}$	$-\left[P_i + \frac{(\frac{1}{\rho} - 1)P_{i|ij} P_{ij} P_{j|ij}}{P_j} \right] \beta_m x_{im}$ if i and j are spatially contiguous $-P_i \beta_m x_{im}$ if i and j are not contiguous
MSCL	$\left\{ \int_{\beta=-\infty}^{\infty} \left[\sum_{j \neq i} (R_{ij}|\beta) \beta_m \right] f(\beta|\theta)\, \mathrm{d}\beta \right\} \frac{x_{im}}{P_i}$, where $R_{ij}|\beta = (P_{i|ij}|\beta)(P_{ij}|\beta)[(1 - P_i|\beta) + \left(\frac{1}{\rho} - 1 \right)(1 - P_{i|ij}|\beta)]$	$\left\{ -\int_{\beta=-\infty}^{\infty} \left[(P_i|\beta)(P_j|\beta) + \left(\frac{1}{\rho} - 1 \right)(P_{i|ij}|\beta) \right. \right.$ $(P_{ij}|\beta)(P_{j|ij}|\beta)]\beta_m f(\beta|\theta)\, \mathrm{d}\beta \right\} \frac{x_{im}}{P_j}$ if i and j are spatially contiguous $\left\{ -\int_{\beta=-\infty}^{\infty} (P_i|\beta)(P_j|\beta)\beta_m f(\beta|\theta)\, \mathrm{d}\beta \right\} \frac{x_{im}}{P_j}$ if i and j are not contiguous

[a] Direct elasticity refers to the percentage change in the choice probability of alternative i due to a 1% change in the mth variable associated with alternative i.

[b] Cross-elasticity is the percentage change in the choice probability of alternative j due to a 1% change in the mth variable associated with alternative i.

2.3. The MSCL model

The previous two sections have discussed the structure and properties of the SCL model. The SCL model maintains the assumption of homogeneity across individuals in the responsiveness to exogenous determinants of residential location choice. However, it is very likely that this responsiveness will vary across individuals, due to observed and unobserved characteristics. The variation in responsiveness due to observed characteristics can be incorporated within the SCL structure by specifying interaction effects of individual-related variables with relevant spatial unit-related attributes. For example, consider the case of location choice decisions for one-worker households. Given the work location, households may choose residential locations based on the sensitivity of the worker to commute time. This sensitivity may be different between men and women, which can be included within the context of the SCL model by adding an interaction variable for commute time and women. However, the variation in responsiveness to commute time due to unobserved factors (or unobserved heterogeneity) cannot be included in the SCL model.

The approach adopted in this study to accommodate unobserved response heterogeneity assumes that the sensitivity variation across individuals can be represented by a continuous distribution; that is, the coefficient vector β embedded in the V_i vector in the SCL model is assumed to be multivariate normal with a vector θ of underlying moment parameters. Let f represent the density function of the multivariate distribution and let F be the corresponding distribution function. Then, the probability of choice of alternative i in the MSCL model may be written as:

$$P_i = \int_{-\infty}^{\infty} (P_i|\beta) f(\beta|\theta)\, \mathrm{d}\beta, \text{ where}$$

$$P_i|\beta = \frac{\sum_{j \neq i} (\alpha_{ij} \mathrm{e}^{\beta' x_i})^{1/\rho} [(\alpha_{i,ij} \mathrm{e}^{\beta' x_i})^{1/\rho} + (\alpha_{j,ij} \mathrm{e}^{\beta' x_j})^{1/\rho}]^{\rho-1}}{\sum_{k=1}^{I-1} \sum_{l=i+1}^{I} [(\alpha_{k,kl} \mathrm{e}^{\beta' x_k})^{1/\rho} + (\alpha_{l,kl} \mathrm{e}^{\beta' x_l})^{1/\rho}]^{\rho}}. \tag{10}$$

The self- and cross-elasticities for the MSCL model are provided in the last row of Table 2. The cross-elasticity expressions in the MSCL model do not exhibit the equal proportional change propensity for non-contiguous alternatives to *i*. This is different from the cross-elasticity expressions in the SCL model.

The dimensionality of the integration in the MSCL probability expression of Eq. (10) is equal to the number of random elements in the coefficient vector β. The reader will note that if an MMNL model were to be used to capture both the spatial autocorrelation in spatial units and unobserved heterogeneity, the dimensionality of the integral in the choice probability would be equal to the number of random elements in the vector β plus the number of paired nests of adjacent alternatives. Even in the very simple case of choice among five spatial units (Fig. 1), this would lead to a multidimensional integral of the order of seven plus the number of random parameters in β. In the empirical context considered in this paper, the dimensionality of the integration would be in the order of 500 if the MMNL model were used. On the other hand, use of the MSCL model reduces the dimensionality to 3.

2.4. Estimation of the MSCL model

The parameters to be estimated in the MSCL model include the scalar ρ representing spatial correlation and the θ vector characterizing the multivariate normal distribution of the β parameters. For ease in presentation, we will absorb the ρ scalar in the parameter vector θ. The estimation of the MSCL model can be pursued using the maximum likelihood method. In particular, the log-likelihood function is:

$$L(\theta) = \sum_{q=1}^{Q} \sum_{i=1}^{I} y_{qi} \log P_{qi}(\theta), \tag{11}$$

where the index *q* has been added to represent households, and

$$y_{qi} = \begin{cases} 1 & \text{if the } q\text{th household chooses spatial unit } i, \\ 0 & \text{otherwise.} \end{cases}$$

The log-likelihood function in Eq. (11) involves the evaluation of multidimensional integrals. In the current study, we apply simulation techniques to approximate the multidimensional integrals and maximize the resulting simulated log-likelihood function. The simulation technique entails computing the integrand in Eq. (11) at several values of the β vector drawn from the multivariate normal distribution for a given value of the parameter vector θ and averaging the integrand values. Notationally, let $\widetilde{P}_{qi}^{n}(\theta)$ be the realization of the choice probability for the *q*th household in the *n*th draw ($n = 1, 2, \ldots, N$). The choice probabilities are then computed as:

$$\widetilde{P}_{qi}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \widetilde{P}_{qi}^{n}(\theta), \tag{12}$$

where $\widetilde{P}_{qi}(\theta)$ is the simulated choice probability of the *q*th household choosing alternative *i* given the parameter vector θ. $\widetilde{P}_{qi}(\theta)$ is an unbiased estimator of the actual probability. Its variance decreases as *N* increases. It also has the appealing properties of being smooth (i.e., twice differentiable) and being positive for any realization of the finite *N* draws.

The simulated log-likelihood function is constructed as:

$$SL(\theta) = \sum_{q=1}^{Q} \sum_{i=1}^{I} y_{qi} \log \widetilde{P}_{qi}(\theta). \tag{13}$$

The parameter vector θ is estimated as the vector value that maximizes the above simulated function. Under rather weak regularity conditions, the maximum simulated log-likelihood estimator is consistent, asymptotically efficient, and asymptotically normal (see Hajivassiliou and Ruud, 1994; Lee, 1992).

In the current paper, we use the Halton sequence to draw realizations for β from its population normal distribution. Details of the Halton sequence and the procedure to generate this sequence are available in Bhat (2002b). Bhat (2002b) has demonstrated that the Halton simulation method out-performs the traditional pseudo-Monte Carlo methods for mixed discrete choice models (see also Hensher, 1999; Train, 1999 for a similar result).

One final note on model estimation. The estimation of the MSCL model includes all the spatial units in the choice set. For the MNL model, a "random sampling of alternatives" approach yields consistent estimates because the correction terms for alternative sampling bias cancel out in the choice probabilities due to the uniform conditioning property of the random sampling strategies (see McFadden, 1978). However, there is no such simplification for non-MNL models.

3. Application of the MSCL model to household residential location choice

3.1. Background

The integrated analysis of land-use and transportation interactions has gained renewed interest and importance with the passage of the Inter-modal Surface Transportation Efficiency Act (IS-TEA) and the Transportation Equity Act for the 21st Century (TEA-21). In this context, one of the most important household decisions is that of residential location, especially because residential land-use occupies about two-thirds of all urban land and home-based trips account for a large proportion of all travel (Harris, 1996). The household residential location decision not only determines the association between the household and the rest of the urban environment, but also influences the household's budgets for activity travel participation.

To be sure, there is a substantial and rich body of literature related to household residential choice. One stream of research on residential location modeling is based on a discrete choice formulation. Sermons and Koppelman (2001) identify at least two appealing characteristics of such a formulation for residential location analysis. First, the discrete choice approach is based on microeconomic random utility theory and models the residential location choice decisions as a trade-off among various locational attributes such as commute time, housing costs, and accessibility to participation in activities. Second, the discrete choice approach allows the sensitivity to locational attributes to vary across sociodemographic segments of the population through the inclusion of interaction variables of locational characteristics with demographic characteristics of households.

The early applications of the discrete choice formulation to residential location analysis include the works of McFadden (1978), Lerman (1975), Onaka and Clark (1983), Weisbrod et al. (1980),

Quagley (1985) and Gabriel and Rosenthal (1989). More recent applications include Timmermans et al. (1992), Hunt et al. (1994), Waddell (1993, 1996), Abraham and Hunt (1997), Ben-Akiva and Bowman (1998), Sermons (2000), and Sermons and Koppelman (2001). Some of the above studies have focused only on residential location choice (for example, McFadden, 1978; Gabriel and Rosenthal, 1989; Weisbrod et al., 1980; Hunt et al., 1994; Sermons and Koppelman, 2001), while others have focused on residential choice as one element of a larger mobility-travel decision-making framework (for example, Lerman, 1975; Quagley, 1985; Waddell, 1993, 1996; Abraham and Hunt, 1997; Ben-Akiva and Bowman, 1998). Similarly some studies have focused on location choice for specific demographic groups (such as single worker and Caucasian households), while others have been more inclusive.

The current study focuses only on residential location choice behavior of households. However, it accommodates measures of accessibility for participation in different purposes as explanatory variables. Thus, the research bears some similarity with earlier studies that have considered residential location within the context of broader mobility and travel decisions. The research is confined to single worker households.

The current research may be distinguished from earlier studies in several respects. First, the research considers spatial autocorrelation in residential location choice decisions. Earlier studies have used a MNL model for residential location choice, which is unable to accommodate spatial autocorrelation. Second, the research uses the full set of spatial alternatives in the estimation process rather than a sample of alternatives. Third, the research considers unobserved variations in sensitivity across individuals to locational and commute-related attributes. Finally, the research considers a reasonably comprehensive set of determinants of residential location choice, based on the empirical findings from earlier studies.

3.2. Data source and sample

The area selected for this study covers part of Dallas County, which is situated in North-Central Texas, and includes the cities of University Park, Highland Park, and Dallas. The area represents 98 out of 383 Transport Analysis and Processing (TAP) zones in Dallas County.

The primary source of data is the 1996 Dallas–Fort Worth (D-FW) metropolitan area household activity survey. This survey collected information about travel and non-travel activities undertaken during a weekday by members of 4839 households, as well as the residential locations of households. The survey also obtained individual and household sociodemographic information. In addition to the activity survey, five other data sets associated with the D–FW metropolitan area were used: land-use/demographic coverage data, the zone-to-zone travel LOS data, school-rating data, census data, and public use microdata sample (PUMS) data. These data are discussed in the next two paragraphs.

Both the land-use/demographic and LOS data files were obtained from the North Central Texas Council of Governments (NCTCOG), which is a voluntary association of local governments from 16 counties in the D–FW urban region. The land-use/demographic data file was used to obtain the total acreage, acreage in specific land-use purposes (including water area, park land, roadways, office area and retail area), the number of households, and the total population for each Transportation Analysis Process (TAP) zone. In addition, the average income and average household size for each TAP zone was computed from this file. The LOS file provided information

on travel between each pair of the 919 TAP zones in the North Central Texas region. The file contains the inter-zonal distances as well as peak and off-peak travel times and costs for transit and highway modes. The land-use/demographic and LOS files, along with data from the activity survey, was used to develop measures of accessibility to activity opportunities, as discussed in the next section.

Data about school ratings was compiled in-house from the 2000 district summary of the Accountability Rating System (ARS) for Texas Public Schools. Each school was ranked as exemplary, recognized, acceptable, or low performing (or unacceptable) based on student drop-out rate, attendance rate, and the percentage of students passing the Texas Assessment of Academic Skills (TAAS) (Texas Education Agency, 2000). Next, the census data was used to compute the ethnic composition of each TAP. Finally, the public use microdata sample (PUMS) data was used to obtain an average value of housing cost for each public use microdata area (PUMA), which was then mapped to the TAP zone level to compute average housing cost for each TAP. The final sample for analysis comprised 236 households with single worker.

3.3. Variable specifications

The six spatial data sources discussed in the previous section provide a rich set of variables for consideration in model specification. The variables may be classified into two broad groups. The first group corresponds to zonal size and attractiveness measures, while the second group corresponds to interactions of sociodemographic characteristics of households with zonal size and attractiveness measures. Each of these broad groups of variables is discussed in turn in the next two sections.

3.3.1. Zonal size and attractiveness variables

The zonal size and attractiveness variables are classified broadly into seven groups: size and density variables, land-use structure measures, zonal demographics and cost variables, commute LOS variables, school quality measures, ethnic composition, and accessibility measures. Table 3 lists the variables considered in each group and the data source used in developing the variables. The development of all variables, except the accessibility measures, is quite straightforward. The accessibility measures are of the Hansen-type (Fotheringham, 1986) and take the following form:

$$A_i^{\text{Rec}} = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{(\text{Park Land Acreage}_j)^{\gamma_{\text{Rec}}}}{(\text{Impedance}_{ij})^{\beta_{\text{Rec}}}} \right),$$

$$A_i^{\text{Ret}} = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{(\text{Number Of Retail Employment}_j)^{\gamma_{\text{Ret}}}}{(\text{Impedance}_{ij})^{\beta_{\text{Ret}}}} \right) \tag{14}$$

and

$$A_i^{\text{Emp}} = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{(\text{Number Of Basic Employment}_j)^{\gamma_{\text{Emp}}}}{(\text{Impedance}_{ij})^{\beta_{\text{Emp}}}} \right),$$

Table 3
Exogenous variables considered in the residential choice models

Variable group	Variable	Data source
Zonal size and density	Log of zonal area in squared miles Log of zonal population Log of number of households in zone Number of households per square mile Number of residents per square mile	Zonal land-use/demographic file
Zonal land-use structure	Percentage of zonal area devoted to single family housing Percentage of zonal area devoted to multi family housing Percentage of zonal area devoted to office space Percentage of zonal area devoted to retail use Percentage of zonal area occupied by water Percentage of zonal area occupied by parks Percentage of zonal area vacant Percentage of zonal area devoted to other use	Zonal land-use/demographic file
Zonal demographics and cost variables	Average household income in zone Average household size in zone Average household cost in zone	Zonal land-use/demographic file and Public Use Microdata Samples (PUMS) file
Commute LOS variables	Commute time between work zone and candidate residential zone Commute distance between work zone and candidate residential zone	LOS file
School quality measures	Dummy variable indicating schools rated as exemplary Dummy variable indicating schools rated as recognized Dummy variable indicating schools rated as acceptable Dummy variable indicating schools rated as low performing	School accountability rating system (ARS)
Ethnic composition	Percentage of Caucasian population Percentage of African American population Percentage of Hispanic population Percentage of other ethnicity	Census
Accessibility measures	Accessibility to sociorecreational opportunities Accessibility to shopping opportunities Accessibility to employment opportunities	Activity survey, zonal land-use/demographic file, and LOS file

where A^{Rec} represents the accessibility to recreational opportunities, A^{Ret} represents the accessibility to shopping opportunities, A^{Emp} represents the accessibility to basic employment opportunities, i is the zone index, and N is the total number of zones in the study region. *Impedance$_{ij}$* is a composite highway auto impedance measure of travel between zone i and zone j. This measure is in effective in-vehicle time units (in minutes) and is expressed as follows:

$$\text{Impedance(in IVTT minutes)} = IVTT + \delta \times OVTT\text{(in minutes)} + \eta \times COST\text{(in cents)} \quad (15)$$

The estimates of δ, η, γ_{Rec}, β_{Rec}, γ_{Ret}, β_{Ret}, γ_{Emp}, and β_{Emp} are obtained using a destination choice model (see Guo and Bhat, 2001 for details of this estimation and the results).

The reader will note that large values of the accessibility measures indicate more opportunities for activities in close proximity of that zone, while small values indicate zones that are spatially isolated from such opportunities.

3.3.2. Interaction of household sociodemographics with zonal characteristics

The motivation for considering interaction effects between household sociodemographics and the zonal size and attractiveness variables is to accommodate the differential sensitivity of households to zone-related attributes. For instance, households with children may be more sensitive to the quality of schools. Similarly, we consider a variable defined as the absolute difference between household income and the median zonal income to test for the presence of income segregation; that is, to test if households locate themselves with other households of similar income level. Several other interaction terms were also considered, including interactions of (a) household size with average zonal household size, (b) sex of worker, ethnicity, and household income with zonal density, (c) household income and sex of worker with commute LOS variables, (d) household income and worker's education status with school quality, (e) worker's ethnicity with the ethnic composition of zones, and (f) ethnicity, worker's education status, and household income with accessibility measures.

We arrived at the final specification based on a systematic process of eliminating variables found to be insignificant in earlier specifications and based on considerations of parsimony in representation. The results of the final specification are presented in Table 4 and discussed in the next section.

3.4. Estimation results

3.4.1. The MNL model results

The MNL model results are presented in the second main column of Table 4. The coefficient on the logarithm of zonal area has the expected positive sign, indicating that households are more likely to locate in larger zones than smaller zones. Households are also more likely to locate in zones with high population density. This may be due to better housing availability at these zones or merely a reflection of population clustering. The only significant zonal land-use structure variable is the percentage of zonal area occupied by multifamily housing units. The parameter on this variable indicates a reluctance to locate in areas with a high percentage of multifamily units. The income dissimilarity measure, captured by the absolute difference between the zonal median income and household income confirms the income segregation phenomenon observed in previous studies (Waddell, 1993). The effect of commute time has the expected negative sign; that is, proximity to the employment location of the worker in the household is an important factor in residential location choice. Interactions of the commute time variable with the sex and the race of the worker were examined to test for the presence of gender disparity in commute patterns and greater spatial mismatch for non-Caucasians compared to Caucasians (see McLafferty and

Table 4
Estimation results of the MNL and MSCL models

Variables	Multinomial logit model		Mixed spatially correlated logit model	
	Parameter	t-statistic	Parameter	t-statistic
Logarithm of zonal area (in mile^2)	0.250	2.776	0.286	3.256
Population density (in 10 persons/mile^2)				
Mean	7.685	4.223	6.987	4.049
Standard deviation[a]	0.000	–	9.358	1.600
Percentage of zonal area occupied by multifamily housing				
Mean	−1.319	−2.063	−3.741	−2.919
Standard deviation[a]	0.000	–	4.541	1.914
Absolute difference between zonal median income and household income (in \$100,000)	−1.270	−2.305	−1.056	−1.762
Commute time (in 100's of minutes)				
Mean	−3.673	−2.200	−4.409	−2.441
Standard deviation[a]	0.000	–	6.504	1.180
Percentage zonal Hispanic population interacted with Hispanic dummy variable	1.235	1.214	1.094	1.127
Work accessibility interacted with African–American household head dummy variable	−2.921	−3.891	−2.329	−3.310
Shopping accessibility	5.809	8.350	5.098	5.759
Dissimilarity parameter[b]	1.000	–	0.358	3.541
Number of observations	236		236	
Log-likelihood at convergence	−1013.43		−1000.93	

[a] The standard deviations are implicitly constrained to 0 in the MNL model.
[b] The dissimilarity parameter is implicitly constrained to 1 in the MNL model. The t-statistic for the dissimilarity parameter in the MSCL model is computed with respect to a value of 1.

Preston, 1992; Blumen, 1994; Turner and Niemeier, 1997; MacDonald, 1999). However, these effects did not turn out to be statistically significant. [2]

The interaction effect of the percentage of Hispanic population with the dummy variable identifying if the head of the household is Hispanic indicates that Hispanic households tend to locate in zones with a high percentage of Hispanic population. This observation of racial segregation may be attributed to one or more of the following factors: (a) racial discrimination in the housing market, (b) differences between racial groups in preferences for neighborhood attributes, or (c) a preference to be with others of the same ethnic background. It is indeed interesting that

---

[2] The reason for the lack of gender disparity may be attributed to the use of one-worker households in the current analysis; many previous commuting studies have examined commute time in the context of male and female workers within the same household to test the "household responsibility hypothesis" (this hypothesis asserts that the female partner commutes less than her male partner to perform more of the household maintenance activities in the joint-household; see Sermons and Koppelman, 2001).

such an effect does not apply to African–American or Caucasian households. Finally, the coefficients on the accessibility measures indicate that (a) African–American households are located in areas with poor work accessibility, and (b) all households prefer locations that offer good accessibility to shopping.

The housing cost and school quality variables were, rather surprisingly, not statistically significant (see Sermons and Koppelman, 2001 for a similar result in their study of residential location in the San Francisco Bay Metropolitan area). The lack of influence of these two variables may be a consequence of the resolution used to represent location in this study. Future work should consider a finer geographic resolution for residential location choice modeling.

3.4.2. The MSCL model results

The MSCL model results in Table 4 are similar to those of the MNL model in terms of the directionality of the mean effect of variables. It is not possible to directly compare the magnitude of the effects of variables from the two models because of difference normalizations of the error term variances (the error term variance is normalized to 1 in the MNL model, but is normalized to a higher value in the MSCL model because of the presence of the random heterogeneity mixing terms). However, a couple of interesting observations may be drawn from the relative magnitudes of variable effects in the two models. First, commute time has a higher (mean) effect in the MSCL model relative to other variables, as can be observed from the higher ratio of the coefficient on commute time to other coefficients. Second, the mean negative effect of the percentage of zonal area occupied by multifamily households in a zone is also higher (relative to other variable effects) in the MSCL model. Clearly, the relative effects of variables are not the same in the two models.

Several variables in the MSCL model were specified to have random coefficients, but only those on zonal population density, the percentage of zonal area in multifamily housing and commute travel time had some statistically significant impact. The results show that, while about 77% of households prefer zones with higher population density, 23% prefer zones with low population density. Similarly, 80% of households prefer zones with lower percentage of zonal area devoted to multifamily housing, while 20% prefer zones with a higher percentage. Finally, the random parameter on commute time suggests that 75% of individuals like to live closer to their work place, while 25% prefer locations farther away from their work place. These results indicate heterogeneity in responsiveness across households.

The dissimilarity parameter of the MSCL model is much smaller than, and significantly different from, 1 (note that the *t*-statistic of the dissimilarity parameter in the table is computed with respect to a value of 1). This result indicates a high level of spatial correlation in residential location choice, which the MNL model fails to recognize.

3.4.3. Elasticity effects and data fit

The MNL and MSCL models imply quite different patterns of inter-alternative competition. To demonstrate the differences, Table 5 presents the disaggregate self- and cross-elasticity values for a randomly selected individual in the sample (the table does not indicate an elasticity effect for accessibility to work because the randomly selected individual is Hispanic, and so the interaction effect of being an African–American and work accessibility does not apply). The numbers in Table 5 indicate the self- and cross-elasticities due to an increase in the variables characterizing the actual chosen zone for the randomly selected individual (of course, any zone can be chosen for

Table 5
Disaggregate elasticity effects

Variables	Multinomial logit model			Mixed spatially correlated logit model		
	Self-elasticity	Cross-elasticity w.r.t. an adjacent zone	Cross-elasticity w.r.t. a non-adjacent zone	Self-elasticity	Cross-elasticity w.r.t. an adjacent zone	Cross-elasticity w.r.t. a non-adjacent zone
Log of zonal area (in mile^2)	1.06019	−0.01512	−0.01512	0.98157	−0.07767	−0.01904
Population density (in 10 persons/mile^2)	0.22775	−0.00325	−0.00325	0.10869	−0.00204	−0.00025
Percentage of zonal area occupied by multifamily housing	−0.04791	0.00068	0.00068	−0.17917	0.01752	0.00030
Absolute difference between zonal median income and household income ($100,000)	−0.00217	0.00003	0.00003	−0.00188	0.00015	0.00004
Commute time (100's of minutes)	−1.28851	0.01837	0.01837	−1.4116	0.10670	0.00235
Percentage zonal Hispanic population interacted with Hispanic dummy variable	0.24650	−0.00351	−0.00351	0.29503	−0.23345	−0.00572
Shopping accessibility	0.53541	−0.00763	−0.00763	0.4908	−0.03884	−0.00952

computing elasticity effects, but we focus on only one zone due to space considerations and presentation ease). The cross-elasticities are computed for a randomly selected zone that is adjacent to the zone whose attributes are changed, as well as for a randomly selected non-adjacent zone.

The MNL cross-elasticities are equal for each variable, reflecting the familiar independence from irrelevant alternatives (IIA) propensity. The cross-elasticities for the MSCL model are different due to (a) the correlation generated between each zone and its neighboring zones in the spatially correlated logit formulation, and (b) the random parameter specification on variables (the latter effect leads to different cross-elasticities even within the group of non-adjacent zones and the group of adjacent zones). Overall, the cross-elasticities of the MSCL model reflect the substantially higher sensitivity between adjacent zones (compared to non-adjacent zones) caused by spatial autocorrelation effects (note the substantially smaller values in the last column relative to the values in the last but one column).

The self-elasticities in both MNL and MSCL models clearly indicate the dominant role played by commute travel time in residential choice modeling. The other important determinants of residential choice include zonal area and accessibility to shopping. As indicated earlier, the effect of commute travel time and the percentage of zonal area occupied by multifamily households is estimated to be higher in the MSCL model relative to the MNL model.

The difference in empirical results between the MNL and MSCL models suggests the need to apply formal statistical tests to determine the structure that is most consistent with the data. The models may be compared using a nested likelihood ratio test (the log-likelihood values at con-

vergence for the two models are provided in the last row of Table 4). The result of such a test leads to the clear rejection of the MNL model; that is, the test provides strong evidence that there is spatial correlation in residence choice and variation in responsiveness across households due to unobserved factors (the likelihood ratio test value is 25 which is larger than the chi-squared statistic with four degrees of freedom at any reasonable level of significance).

4. Summary and conclusions

This paper has proposed a MSCL model for the analysis of location-related decisions of individuals and households. The paper submits, and demonstrates, that while the MMNL class of models is very general in structure, these are substantial computational efficiency gains to be achieved by using MGEV structures in spatially correlated choice situations. This is because the number of error components that needs to be specified in the MMNL structure to generate the desired spatial correlation pattern is very high for realistic location choice decisions. This leads to a high dimensionality of integration in the MMNL structure. In the empirical setting of the current paper, the use of a MMNL structure would entail a multidimensional integral of the order of 500, while the proposed MSCL model, which is based on a MGEV structure, requires evaluation of only a three-dimensional integral.

In addition to computational efficiency gains, there is another more basic reason to prefer the MSCL model over an MMNL structure. 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, 2002, p. 191). In this regard, superimposing a mixing structure to accommodate random coefficients, over a closed form analytic structure that accommodates a particular desired inter-alternative error correlation structure, represents a powerful approach to capture random taste variations and complex substitution patterns.

A broad purpose of this paper is to demonstrate that 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. Hopefully, this paper demonstrates that 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 MNL structure.

The empirical analysis in the paper applies the MSCL model to examine the residential choice behavior of households in Dallas County using the 1996 Dallas–Fort Worth metropolitan area household activity survey. The empirical results indicate the important and dominant effect of commute travel time on residential location choice. Other variables significantly impacting residential choice include zone size, population density, percentage of zonal area occupied by multi-family housing, disparity between household income and median zonal income, percentage of Hispanic population for Hispanic households, and work and shopping accessibility.

A comparison of the MNL and MSCL models estimated in the paper indicates the significant presence of spatial correlation between contiguous zonal alternatives as well as differential responsiveness to exogenous variables across households (for example, the results suggest that, while 75% of households prefer living close to their workplace, about 25% prefer to live further away from their workplace). The MSCL model also leads to a statistically superior data fit. In addition, the results indicate that failing to accommodate spatial correlation and unobserved response heterogeneity can lead to incorrect conclusions regarding the elasticity effects of exogenous variables.

The model estimated in the paper can be used to examine the impacts of changes in sociodemographics, transportation LOS, and land-use characteristics on residential location choice. For example, the percentage of Hispanic households in the population is projected to grow from 12.4 to 15.1 in the next decade (U.S. Census Bureau, 2000). The results of our model suggest that these households are likely to cluster around existing Hispanic "centers" and the consequent changes in residential location patterns can be forecasted using the residential choice model developed in the paper. Similarly transportation system changes that affect commute travel time can be examined by the model, and land-use changes can be analyzed by appropriately modifying the density, land-use structure, and accessibility variables.

Acknowledgements

References

Abraham, J.E., Hunt, J.D., 1997. Specification and estimation of a nested logit model of home, workplace and commuter mode choice by multiple worker households. Transportation Research Record 1606, 17–24.

Ben-Akiva, M., Bowman, J.L., 1998. Integration of an activity-based model system and a residential location model. Urban Studies 35 (7), 1131–1153.

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

Ben-Akiva, M., Lerman, S., 1985. Discrete-Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

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

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

Bhat, C.R., 2002a. Recent methodological advances relevant to activity and travel behavior. In: Mahmassani, H.S. (Ed.), Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges. Elsevier Science, Oxford, UK, pp. 381–414.

Bhat, C.R., 2002b. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research B, forthcoming.

Bierlaire, M., 2002. The network GEV model. In: Proceedings of the 2nd Swiss Transportation Research Conference. Monte Verita, Switzerland.

Blumen, R., 1994. Gender differences in the journey to work. Urban Geography 15, 223–245.

Bolduc, D., 1992. Generalized autoregressive errors in the multinomial probit model. Transportation Research B 26 (2), 155–170.

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

Chu, C., 1989. A paired combinatorial logit model for travel analysis. In: Proceedings of the Fifth World Conference on Transportation Research Ventura, CA, pp. 295–309.

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

Fotheringham, A.S., 1986. Modeling hierarchical destination choice. Environment and Planning 18A, 401–418.

Gabriel, S.A., Rosenthal, S.S., 1989. Household location and race: estimates of a multinomial logit model. The Review of Economics and Statistics 17 (2), 240–249.

Garrido, R.A., Mahmassani, H.S., 2000. Forecasting freight transportation demand with the space time multinomial probit model. Transportation Research B 34, 403–418.

Guo, J.Y., Bhat, C.R., 2001. Residential location choice modeling: a multinomial logit approach. Technical paper, Department of Civil Engineering, University of Texas at Austin.

Hajivassiliou, V.A., Ruud, P.A., 1994. Classical estimation methods for LDV models using simulations. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics IV. Elsevier, New York, pp. 2383–2441.

Harris, B., 1996. Land use models in transportation planning: a review of past developments and current practice. Available at: <http://www.bts.gov/other/MFD_tmip/papers/landuse/compendium/dvrpc_appb.htm>.

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

Hensher, D.A., Green, W.H., 2001. The mixed logit model: the state of practice and warnings for the unwary. Working paper, Institute of Transport Studies, University of Sydney.

Hunt, J.D., McMillan, J.D.P., Abraham, J.E., 1994. Stated preference investigation of influences on attractiveness of residential locations. Transportation Research Record 1466, 79–87.

Johnson, N., Kotz, S., 1970. Distribution in Statistics: Continuous Univariate Distributions. John Wiley, New York, Chapter 21.

Koppelman, F.S., Wen, C.-H., 2000. The paired combinatorial logit model: properties, estimation and application. Transportation Research B 34 (2), 75–89.

Lee, L.-F., 1992. On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econometric Theory 8, 518–552.

Lerman, S.R., 1975. A disaggregate behavioral model of urban mobility decisions, Ph.D. dissertation, Massachusetts Institute of Technology.

Luce, R.D., Suppes, P., 1965. Preference utility and subjective probability. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), Handbook of Mathematical Psychology. John Wiley and Sons, New York.

MacDonald, H.I., 1999. Women's employment and commuting: explaining the links. Journal of Planning Literature 13 (3), 267–283.

McFadden, D., 1978. Modeling the choice of residential location. Transportation Research Record 672, 72–77.

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

McLafferty, S., Preston, V., 1992. Spatial mismatch and labor marker segmentation for African American and Latina women. Economic Geography 68 (4), 406–431.

Onaka, J., Clark, W.A.V., 1983. A disaggregate model of residential mobility and housing choice. Geographical Analysis 19, 287–304.

Quagley, J.M., 1985. Consumer choice of dwelling, neighborhood and public services. Regional Science and Urban Economics 15, 41–63.

Sermons, M.W., 2000. Influence of race on household residential utility. Geographical Analysis 32 (3), 225–246.

Sermons, M.W., Koppelman, F.S., 2001. Representing the differences between female and male commute behavior in residential location choice models. Journal of Transport Geography 9, 101–110.

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

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

Texas Education Agency, 2000. 2000 Accountability Manual.

Timmermans, H., Borgers, A., Dijk, J., Oppewal, H., 1992. Residential choice behavior of dual earner households: a decompositional joint choice model. Environment and Planning A 24, 517–533.

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

Train, K., 2002. Discrete Choice Methods with Simulation. Cambridge University Press, New York.

Turner, T., Niemeier, D., 1997. Travel to work and household responsibility: new evidence. Transportation 24, 397–419.

U.S. Census Bureau, 2000. National Population Projections. Available at: <http://www.census.gov/population/www/projections/natdet.html>.

Vovsha, P., 1997. The cross-nested logit model: application to ode choice in the Tel-Aviv metropolitan area. Transportation Research Record 1607, 6–15.

Waddell, P., 1993. Exogenous workplace choice in residential location models: is the assumption valid? Geographical Analysis 25, 65–82.

Waddell, P., 1996. Accessibility and residential location: the interaction of workplace, residential mobility, tenure, and location choices, presented at the Lincoln Land Institute TRED Conference. Available at: <http://www.odot.state.or.us/tddtpan/modeling.html>.

Weisbrod, G., Lerman, S., Ben-Akiva, M., 1980. Tradeoffs in residential location decisions: transportation vs. other factors.. Transportation Policy and Decision Making 1, 13–26.

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

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