

**ACCOMMODATING SPATIAL CORRELATION ACROSS CHOICE ALTERNATIVES
IN DISCRETE CHOICE MODELS: AN APPLICATION TO MODELING
RESIDENTIAL LOCATION CHOICE BEHAVIOR**

Ipek N. Sener

Department of Civil, Architectural and Environmental Engineering
1 University Station, C1761
The University of Texas at Austin
Austin, Texas 78712
Phone: (512) 471-4535; Fax: (512) 475-8744
Email: ipek@mail.utexas.edu

Ram M. Pendyala

Department of Civil and Environmental Engineering
PO Box 875306, ECG252
Arizona State University
Tempe, AZ 85287-5306
Phone: (480) 727-9164; Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Chandra R. Bhat (*corresponding author*)

Department of Civil, Architectural and Environmental Engineering
1 University Station, C1761
The University of Texas at Austin
Austin, Texas 78712
Phone: (512) 471-4535; Fax: (512) 475-8744
Email: bhat@mail.utexas.edu

ABSTRACT

This paper presents a modeling methodology capable of accounting for spatial correlation across choice alternatives in discrete choice modeling applications. Many location choice (*e.g.*, residential location, workplace location, destination location) modeling contexts involve choice sets where alternatives are spatially correlated with one another. In the presence of spatial correlation, traditional discrete choice modeling methods that are often based on the assumption of independence among choice alternatives are not appropriate. Previous work in the field has yielded methods that can accommodate correlation among neighboring or adjacent location choices, but has not provided the ability to accommodate spatial correlations that may exist across choice alternatives that are not necessarily neighbors or adjoining one another. In this paper, a generalized spatially correlated logit (GSCL) model, and its mixed version, that allows one to represent the degree of spatial correlation as a function of a multidimensional vector of attributes characterizing each pair of location choice alternatives is formulated and presented. The model is applied to the analysis of residential location choice behavior using a sample of households drawn from the 2000 San Francisco Bay Area Travel Survey (BATS) data set. Model estimation results obtained from the GSCL are compared against those obtained using the standard multinomial logit (MNL) model and the spatially correlated logit (SCL) model where only correlations across neighboring (or adjacent) alternatives are accommodated. Model findings suggest that there is significant spatial correlation across alternatives that do not share a common boundary, and that the GSCL offers the ability to more accurately capture spatial location choice behavior.

Keywords: spatial correlation, spatially correlated logit model, residential location choice, distance-decay function, activity-travel behavior modeling, discrete choice modeling

1. INTRODUCTION

Emerging activity-based approaches to travel analysis explicitly recognize interactions among activities, trips, and individuals in time and space. Many choices encountered in land use and transportation modeling are *spatial* in nature. Individuals make decisions about where to live and work, where to go to school, where to pursue various activities such as shopping, personal business, and social-recreation, and which route to take when traveling between an origin-destination pair. Despite the clear recognition of the importance of the space dimension in modeling people's location choice behavior, research advances in modeling spatial effects in the travel behavior field has generally lagged advances in modeling temporal effects. There has been much research in understanding time use patterns, modeling temporal constraints associated with time-space prisms, and analyzing trade-offs and synergies in time allocation across activities of various types and across individuals in a household.

The modeling of spatial effects in activity-travel behavior analysis has generally lagged that of its temporal counterpart for two main reasons. First, modeling spatial dependencies and interactions is inherently more complex due to the difficulty in characterizing, defining, and measuring such effects. Second, activity-travel behavior data sets offer rich temporal information, but often lack much detail along the spatial dimension. Location information in typical activity-travel data sets tend to be either missing entirely, or where available, is coarse in nature – either because of the difficulty in measuring spatial dimensions or because of concerns regarding respondent privacy. However, in recent years, advances have been made on both fronts. Data sets from activity-based travel surveys are beginning to offer richer information about spatial choices and the availability of detailed spatial information is only going to get better with the increasing use of GPS-based travel surveys. Second, advances in discrete choice modeling, both in terms of formulation and estimation, provide the framework for incorporating complex spatial effects in models of activity-travel behavior.

This paper addresses a key spatial effect known as spatial autocorrelation or spatial correlation across alternatives. Spatial correlation across alternatives is prevalent when location alternatives in a choice set are correlated or related to one another. When alternatives share common observed and unobserved attributes that influence choice-maker behavior, then the fundamental assumptions of independence across choice alternatives that form the basis of the multinomial logit formulation are violated. In virtually any location choice context, one would

expect alternatives (locations) that are closer to one another to be more correlated with one another than those that are farther apart. In other words, the consideration of proximity is an important criterion in understanding and modeling spatial correlation effects in models of location choice.

In this context, this paper presents a discrete choice modeling methodology that explicitly incorporates spatial correlation across location choice alternatives. The key feature of the proposed modeling methodology is that the extent of spatial correlation is a function of a multidimensional vector of attributes characterizing the relationship between each pair of locations. For instance, distance can be one element of this multi-dimensional vector, since locations that are closer to one another are likely to be more correlated than others. Similarly, whether or not locations share a common boundary and the length of the shared common boundary can also be elements of the multidimensional vector. Such a formulation constitutes a significant advancement over previous attempts to represent spatial correlation where typically only those locations that share a common boundary (*i.e.*, locations are immediate neighbors and adjacent to one another) are considered to have a non-zero correlation and all other location pairs are assumed to be spatially uncorrelated. By incorporating generalized spatial correlation patterns into advanced discrete choice modeling methods, one can model an array of discrete behavioral phenomena while accommodating flexible substitution patterns in a random utility-maximization framework.

The Generalized Spatially Correlated Logit (GSCL) model formulated and presented in this paper is applied to the context of residential location choice analysis, an important choice dimension that influences and is influenced by the built environment and human activity-travel patterns. Virtually all integrated land use – transportation microsimulation model systems include residential location models as a key component of the framework. As many activity-travel choices are influenced by built environment attributes associated with residential locations, it is of utmost importance and interest to ensure that the residential location choice model component is robust with respect to accounting for spatial correlation effects that inevitably exist in residential location choice contexts. When considering residential location choice alternatives, individuals often consider agglomerations of zones or spatial units that are not only adjacent to one another, but also closer to one another. This is likely due to the common observed and unobserved attributes that spatial units in a “proximal cluster” share with one another with

respect to socio-demographic composition of residents, income levels, density and pattern of development, proximity to services and shopping opportunities, availability of green space, and transit and pedestrian/bicycle-friendliness.

The GSCL model is applied in this paper to examine residential location choices of a sample of households drawn from the 2000 San Francisco Bay Area Travel Survey (BATS) data set. The activity-based travel survey data set is augmented with a host of secondary built environment variables to facilitate the specification and estimation of a residential location choice model that incorporates spatial dependencies across location alternatives. The traffic analysis zone (TAZ) is treated as the spatial unit of analysis as most activity-travel demand models continue to use zone-level data sets, presumably because secondary data is available at this level of spatial aggregation.

The remainder of this paper is organized as follows. The next section presents a detailed description of spatial effects that are worthy of consideration and how these effects have been accommodated in models in past work. The third section presents a detailed formulation of the Generalized Spatially Correlated Logit (GSCL) model proposed in this paper. The fourth section describes the data set used for the residential model application while the fifth section presents detailed model estimation results. Conclusions are presented in the sixth and final section.

2. UNDERSTANDING AND REPRESENTING SPATIAL EFFECTS IN CHOICE MODELS

2.1 Analysis Context

This paper focuses on accounting for spatial effects in discrete choice behaviors of interest in an activity-travel context. In such a context, spatial effects refer to the dependence (or correlation) and variation (heterogeneity) of activity-travel behavior across space (Hunt *et al.*, 2004; Páez *et al.*, 2008). Both spatial correlation and spatial heterogeneity may occur across choice alternatives or across decision-makers and it is important to explicitly incorporate consideration of such spatial effects in activity-based travel demand model systems that aim to capture time-space interactions underlying activity-travel patterns in a microsimulation modeling framework.

The paper specifically recognizes and addresses the issue of spatial correlation across choice alternatives. The existence of such correlation is best motivated by considering the First Law of Geography, as suggested by Tobler (1970): “Everything is related to everything else, but

near things are more related than distant things”. Spatial correlation among choice alternatives arises due to common unobserved locational effects that influence the choice behavior of decision makers (Guo, 2004). For example, there may be a perceived similarity between neighboring or adjoining spatial choice alternatives as opposed to those that are farther apart (Bhat and Zhao, 2002). Although this paper focuses on spatial correlation across location alternatives in the choice set, it should be recognized that spatial correlation may also exist across choice-makers, an issue discussed at length elsewhere (see Fleming, 2004; and Bhat and Sener, 2008).

2.2 Overview of Earlier Relevant Research

It has been long recognized that, while the multinomial logit (MNL) model offers computational tractability even in the presence of large choices sets, it suffers from potential violations of the IIA property arising from correlated choice alternatives that can lead to inconsistent parameter estimates and unrealistic forecasts (Horowitz, 1981; Train, 2003). Hunt *et al.* (2004) and Haynes and Fotheringham (1990) have indicated that the IIA property is unlikely to hold true in spatial choice applications where alternatives are characterized by size, dimensionality, aggregation, location characteristics, and spatial continuity and variation. Unlike the MNL, the nested logit (NL) model (see Williams, 1977; Daly and Zachary, 1978; McFadden, 1978) assumes a hierarchical choice structure, which makes it appropriate for representing various spatial choice behaviors, while potentially avoiding violations of the IIA property (see, for example, Waddell, 1996; Abraham and Hunt, 1997; Boots and Kanaroglou, 1988; Deng *et al.*, 2003). However, the NL model is not without its limitations. As noted by Pellegrini and Fotheringham (2002), the members of each cluster or nest of alternatives must be specified a priori in the NL model, requiring that space be divided meaningfully without accommodating the full range of spatial substitutability that may occur.

Several researchers have indeed formulated and estimated more advanced discrete choice models than the MNL and NL models to incorporate spatial correlation. For example, Bolduc *et al.* (1996) uses a mixed logit model formulation that adopts a first-order spatial autoregressive process and apply this framework to model the initial practice location (among 18 spatial alternatives) for general medical practitioners. Garrido and Mahmassani (2000) propose a spatially (and temporally) correlated multinomial probit model for analyzing the probability that

a shipment of a particular commodity will originate in a certain spatial unit at a certain time interval of the day. Their model choice set includes between 31 and 41 location alternatives. Miyamoto *et al.* (2004) use a framework similar to Bolduc *et al.* (1996) for the error autocorrelation, but also include an autocorrelated deterministic component of utility. They estimate a mixed logit structure using residential location choice data for four specific zones in the city of Sendai in Japan. All of these studies employed simulation-based techniques for model estimation due to the open-form nature of the choice probabilities that entail evaluation of a J -dimensional integral in the likelihood function where J is the number of choice alternatives. However, even with recent advances in simulation approaches, model estimation becomes prohibitive and potentially affected by simulation error in the presence of large choice sets.

An important development in the field of discrete choice modeling was the introduction of the Generalized Extreme Value (GEV) class of models within the random utility maximization framework (McFadden, 1978, 1981). The GEV-class of models allows flexible substitution patterns between different choice alternatives, while maintaining a simple closed-form structure for the choice probabilities. Several models have been developed within the GEV class, as recently discussed by Daly and Bierlaire (2006), Koppelman and Sethi (2008), and Bekhor and Prashker (2008). The flexibility of GEV structures also allows the modeling of spatial location choice problems where the utilities of various location choice alternatives may be correlated with one another due to common unobserved spatial elements. Thus, Bhat and Guo (2004) proposed a GEV-based model formulation called the *spatially correlated logit* (SCL) model that results in a closed-form expression regardless of the number of alternatives in the choice set. They apply the SCL model to analyze residential location choice behavior among 98 zones in Dallas County in Texas.

The main limitation of the SCL model is that it accommodates spatial correlation only across contiguous location alternatives that share a common border. If the location alternatives are not adjacent to or adjoining one another, then the spatial correlation is assumed to be zero. Although this specification may be applicable in some choice situations, it is likely to prove unnecessarily restrictive in model specifications for most choice situations. In most choice contexts, one expects the degree of spatial correlation to be greater among alternatives that are close to one another and less for those that are farther apart. In other words, one expects that, while there may be a certain level of heightened correlation between adjacent spatial alternatives,

there is also likely to be a decay function-based correlation that decreases as the degree of spatial separation between alternatives increases. This paper is aimed at enhancing the SCL model to accommodate such flexible spatial correlation specifications, while retaining the appealing closed form nature of the SCL model. Random taste preferences due to unobserved decision-maker characteristics can also be captured by introducing a mixed version of the GSCL model.

3. GENERALIZED SPATIALLY CORRELATED LOGIT (GSCL) MODEL FORMULATION

The SCL model proposed by Bhat and Guo (2004) constitutes a GEV-based formulation adopted in the context of spatially correlated discrete choice models. The SCL model is based on a special case of the generalized nested logit (GNL) model originally proposed by Wen and Koppelman (2001).

The SCL model is derived from the following generator function:

$$G(e^{V_{n1}}, \dots, e^{V_{nl}}) = \sum_{i=1}^{I-1} \sum_{j=i+1}^I ((\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu})^\mu \quad (1)$$

In the above formulation, $\alpha_{i,ij}$ represents the allocation parameter that characterizes the portion of alternative i that belongs to the i - j location pair, $0 < \alpha_{i,ij} < 1$ for all i and j , and $\sum_j \alpha_{ij} = 1$ for all i . V_{ni} represents the deterministic component of the utility associated with location alternative i for individual n . μ is a dissimilarity parameter capturing the correlation between spatial units, where $0 < \mu \leq 1$. The SCL model is built upon the assumption that the dissimilarity parameter, μ , is the same for all paired nests. This assumption significantly increases the efficiency of model estimation, especially in the case of spatial choice models where the number of alternatives (or locations) is large. As indicated in Bhat and Guo (2004), the correlation between two alternatives in the generator function form of Equation (1) increases as μ gets closer to zero, and also increases as $\alpha_{i,ij}$ and $\alpha_{j,ij}$ increase (the correlation values need to be computed numerically, and are tabulated in Bhat and Guo for various combinations of μ ,

$\alpha_{i,ij}$ and $\alpha_{j,ij}$).¹ When $\mu = 1$, the generator function in Equation (1) collapses to that of the MNL model generator function, and the SCL model becomes equivalent to the MNL model.

The probability function of an individual n choosing alternative i in the SCL model is as follows:

$$P_{ni} = \sum_{j \neq i} \left(\frac{(\alpha_{i,ij} e^{V_{ni}})^{1/\mu}}{(\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu}} \times \frac{((\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu})^\mu}{\sum_{k=1}^{I-1} \sum_{l=k+1}^I ((\alpha_{k,kl} e^{V_{nk}})^{1/\mu} + (\alpha_{l,kl} e^{V_{nl}})^{1/\mu})^\mu} \right) \quad (2)$$

The SCL model is formulated so that only adjacent (or contiguous) alternatives share common unobserved characteristics, which is consistent with the way spatial correlation has been captured in much of the spatial econometrics literature. Specifically, in the SCL model, the allocation parameter, $\alpha_{i,ij}$, is defined as the proportion of number of neighboring zonal alternatives. This formulation is consistent with the assumption that sensitivity to changes in neighboring spatial units is larger for a zone with fewer neighboring zones. Specifically, in the SCL model, the allocation parameter for zone i is defined as follows:

$$\alpha_{i,ij} = \frac{\omega_{ij}}{\sum_k \omega_{ik}} \quad (3)$$

In the above allocation equation, ω_{ij} is 1 if zone i is contiguous to zone j , and 0 otherwise (by convention $\omega_{ii} = 0$ for all i). The specification of the allocation parameter as in Equation (3) implies equal allocation of zone i to each nest formed by pairing i with each of its adjacent alternatives. Thus, zone i is equally correlated with all neighboring zones that themselves have the same number of adjacent zones to them.

One apparent problem with the contiguity assumption adopted in the SCL model formulation is the widely recognized modifiable areal unit problem (MAUP) (Openshaw, 1984). The presence of MAUP results from the use of aggregate geographical data, such as census tracts or zones, when micro-data is not available. In particular, MAUP results in model estimates being sensitive to the zonal configuration. Any change in zonal configuration leads to a critical problem, especially in spatial location choice modeling contexts where contiguity is taken as the key parameter to account for the correlation among spatial choice alternatives. As Fotheringham

¹ See also Marzano and Papola (2008) and Abbe *et al.* (2007) for discussions of the correlation structure for the GNL model.

and Brunson (1999) note “[b]y changing the zonal system, not only will the aggregated data alter, but also the points in the study area that are considered to be in adjacent zones”. In addition to being subject rather severely to the MAUP, the SCL does not accommodate spatial correlation across alternatives that are close, but not contiguous, to one another, as discussed earlier. In this respect, the model proposed in this paper extends the SCL model in an important way by accommodating spatial correlation across *all* choice alternatives as opposed to a strictly contiguity-based formulation.

Spatial correlation across all choice alternatives is accommodated by introducing an allocation parameter which is a function of a multidimensional vector of attributes characterizing the relationship between spatial choice alternatives. The only requirement is that this mapping should continue to satisfy the following two conditions for the allocation parameters: $0 < \alpha_{i,j} < 1$ for all i and j , and $\sum_j \alpha_{ij} = 1$ for all i . While many functional forms can be used for this mapping, we use the simple multinomial logit form in the proposed generalized spatially correlated logit (GSCL) model:

$$\alpha_{i,j} = \frac{\exp(\phi'z_{ij})}{\sum_k \exp(\phi'z_{ik})} \quad (4)$$

where z_{ij} is a vector of attributes characterizing the spatial relationship between alternatives i and j (by convention, $z_{ii} = 0$ for all i), and ϕ is a corresponding coefficient vector to be estimated. The z_{ij} vector may include variables such as the shared boundary length for contiguous alternatives, the inter-alternative distance, whether alternatives share some common observed characteristic (such as say presence of a mall or presence of a large park or being part of the same “neighborhood”), similarity measures based on observed location characteristics (such as intensity or density of land-use in certain types of activities), presence of physical barriers such as rivers and mountains, and other measures of network connectivity and topography.²

² The use of the functional form in Equation (4) allows relatively easy interpretation of the ϕ coefficient vector. Thus, for example, if the element of the ϕ vector corresponding to shared boundary length is positive, it implies that $\alpha_{i,j}$ will be higher than $\alpha_{i,k}$ for a spatial unit j closer to i than is spatial unit k . This has the effect of a higher correlation between alternatives i and j than i and k , as discussed in Section 3.

The reader will note that Equation (3) for the allocation parameters in the SCL model is not a restricted version of Equation (4) for the allocation parameters in the GSCL model. Thus, the GSCL model does not nest the SCL model. However, the GSCL model does generalize the SCL model by allowing multiple variables in the z_{ij} vector, including the contiguity dummy variable ω_{ij} . The GSCL and the SCL models can be formally compared using a non-nested likelihood ratio test. On the other hand, the GSCL model, like the SCL model, nests the MNL model, which is obtained when $\mu = 1$.³

One other special case of the GSCL model that we will discuss here (because this was the form that came out to be relevant in the empirical context of the current study) is the distance-based spatially correlated logit (DSCL) model. In particular, consider the case where the only variable in the z_{ij} vector of Equation (4) is inter-alternative distance. While many different functional forms can be tested for the inter-alternative distance (including linear distance, logarithm of distance, “cliff-distance” specifications where zones that exceed a certain distance threshold fall off the correlation “cliff” and become uncorrelated, *etc.*), consider the simple logarithm form of distance in the z_{ij} vector. In this case, Equation (4) collapses to the following form:

$$\alpha_{i,ij} = \frac{\exp(\phi' \ln d_{ij})}{\sum_k \exp(\phi' \ln d_{ik})} = \frac{d_{ij}^\phi}{\sum_k d_{ik}^\phi} \quad (5)$$

The expectation in the above DSCL model allocation form is that the parameter ϕ is negative, so that correlation between alternatives reduces as the distance between them increases. This is essentially a distance decay function for the allocation parameters.

The GSCL model is derived based on the following generator function:

$$G(e^{V_{n1}}, \dots, e^{V_{nl}}) = \sum_{i=1}^{I-1} \sum_{j=i+1}^I \left((\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu} \right)^\mu \quad (6)$$

where $\alpha_{i,ij}$ as specified in Equation (4). Based on the function given in Equation (6), the vector of the unobserved portion of utility, $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2}, \dots, \varepsilon_{nl})$, has the following cumulative extreme-value distribution:

³ When $\mu = 1$ in the GSCL model, the ϕ parameter vector in Equation (4) becomes econometrically unidentifiable.

$$F(\varepsilon_n) = \exp \left\{ - \sum_{i=1}^I \sum_{j=i+1}^{I-1} \left[(\alpha_{i,ij} e^{-\varepsilon_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{-\varepsilon_{nj}})^{1/\mu} \right]^\mu \right\} \quad (7)$$

Note that the marginal cumulative distribution function (CDF) of each stochastic element ε_i has a univariate extreme-value distribution, which is the standard Gumbel distribution function:

$$F(\varepsilon_i) = \exp \left\{ - \sum_{j \neq i} \alpha_{i,ij} e^{-\varepsilon_i} \right\} = \exp \left\{ - e^{-\varepsilon_i} \right\} \quad (8)$$

The bivariate marginal CDF for two stochastic elements ε_i and ε_k of two spatial units i and k , whether i and k are adjacent or not, is as follows:

$$H(\varepsilon_i, \varepsilon_k) = \exp \left\{ - (1 - \alpha_{i,ik}) e^{-\varepsilon_i} - (1 - \alpha_{k,ik}) e^{-\varepsilon_k} - \left[(\alpha_{i,ik} e^{-\varepsilon_i})^{1/\mu} + (\alpha_{k,ik} e^{-\varepsilon_k})^{1/\mu} \right]^\mu \right\} \quad (9)$$

The probability function of the GSCL model (*i.e.* the probability of decision maker n choosing alternative i) is obtained by substituting the generator function of Equation (6) into McFadden's (1978) probability expression:

$$\begin{aligned} P_{ni} &= \frac{\sum_{j \neq i} (\alpha_{i,ij} e^{V_{ni}})^{1/\mu} \left((\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu} \right)^{\mu-1}}{\sum_{k=1}^{I-1} \sum_{l=k+1}^I \left((\alpha_{k,kl} e^{V_{nk}})^{1/\mu} + (\alpha_{l,kl} e^{V_{nl}})^{1/\mu} \right)^\mu} \\ &= \sum_{j \neq i} \left(\frac{(\alpha_{i,ij} e^{V_{ni}})^{1/\mu}}{(\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu}} \times \frac{\left((\alpha_{i,ij} e^{V_{ni}})^{1/\mu} + (\alpha_{j,ij} e^{V_{nj}})^{1/\mu} \right)^\mu}{\sum_{k=1}^{I-1} \sum_{l=k+1}^I \left((\alpha_{k,kl} e^{V_{nk}})^{1/\mu} + (\alpha_{l,kl} e^{V_{nl}})^{1/\mu} \right)^\mu} \right) \\ &= \sum_{j \neq i} P_{i|ij} P_{ij}. \end{aligned} \quad (10)$$

The direct and cross elasticity formulas for the GSCL model are similar to those derived for the SCL model, except for the cross elasticity expressions for non-adjacent alternatives (see Table 1 for the elasticity expressions). As can be seen from the elasticity formulations, the cross elasticities for the SCL model are higher than the MNL model only for spatial units adjacent to alternative i . In other words, the cross elasticity expression of the SCL model is the same as that for the standard MNL for all choice alternatives not adjacent to alternative i . This is because the correlation between alternative i and all non-adjacent alternatives is assumed to be zero in the SCL, which is consistent with the assumption of independence of choice alternatives intrinsic to

the standard MNL formulation. On the other hand, the GSCL model relaxes this assumption and does not yield constant cross elasticities, even for alternatives that are not adjacent to alternative i because correlation between non-adjacent locations is accommodated.

The GSCL model discussed thus far accommodates correlation across spatial units regardless of adjacency. However, it does not consider heterogeneity across decision makers in the responsiveness to exogenous determinants of the choice alternatives. This sensitivity variation can be incorporated by superimposing a mixing distribution into the GEV structure of the GSCL model. Such unobserved response heterogeneity can be accommodated by representing the distribution of the unobserved heterogeneity with a continuous function. In particular, the coefficient vector β embedded in the V_{ni} vector of the GSCL model may be assumed to be multivariate normal with a vector θ of underlying moment parameters. Then, the probability of choosing alternative i by decision maker n in the mixed GSCL (MGSCl) model is given as follows:

$$P_{ni} = \int_{-\infty}^{\infty} (P_{ni} | \beta) f(\beta | \theta) d\beta, \quad (11)$$

where

$$P_{ni} | \beta = \frac{\sum_{j \neq i} (\alpha_{i,j} e^{\beta' x_{ni}})^{1/\mu} \left[(\alpha_{i,i,j} e^{\beta' x_{ni}})^{1/\mu} + (\alpha_{j,i,j} e^{\beta' x_{nj}})^{1/\mu} \right]^{\mu-1}}{\sum_{k=1}^{I-1} \sum_{l=k+1}^I \left[(\alpha_{k,kl} e^{\beta' x_{nk}})^{1/\mu} + (\alpha_{l,kl} e^{\beta' x_{nl}})^{1/\mu} \right]^{\mu-1}}, \quad (12)$$

and f represents the density function of the multivariate normal distribution.

The parameters to be estimated in Equation (12) include the β vector representing the observed parameters for the choice attributes, the scalar dissimilarity parameter μ , the coefficient vector ϕ embedded in the allocation parameters, and the θ vector characterizing the multivariate normal distribution of the β parameters. Define $m_{ni} = 1$ if decision maker n chooses the spatial unit i , and 0 otherwise. Then, the likelihood function for estimation of the various parameters collected into a single vector γ can be written as:

$$L(\gamma) = \sum_{n=1}^N \sum_{i=1}^I m_{ni} \ln P_{ni}(\gamma). \quad (13)$$

The evaluation of the log-likelihood function in Equation (13) requires the computation of the multivariate integral in the choice probability whose dimensionality is equal to the number of

random elements in the coefficient vector β . It should be noted that, if the mixed logit model is used instead of the MGSCCL model, 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 all paired nests. Thus, the GEV-based formulation underlying the GSCL model substantially increases the computational efficiency even if mixing is introduced. This efficiency gain is especially important in spatial location choice models where the number of paired nests is very large. Simulation techniques can be applied to approximate the multidimensional integrals in the likelihood function and maximize the resulting simulated log-likelihood function.

To summarize, there are at least two key benefits of using the generalized allocation (GSCL) form of Equation (4) rather than the strictly contiguity-based allocation (SCL) form of Equation (3). First, the generalized formulation reduces the impact of the MAUP on model estimates, because a zone pair that shares a common boundary in one spatial configuration could become non-adjacent zones in a different spatial configuration. The correlation between this pair of zones would instantaneously change from a positive value to zero in the SCL model as a result of the change in zonal configuration. On the other hand, the impact of zonal configuration on model estimates is both accommodated and moderated in the GSCL formulation. Second, the GSCL model relaxes the assumption of zero allocation, and hence correlation, across non-contiguous spatial units. By taking advantage of the introduction of a general function in the formulation of the allocation parameter, the GSCL model is ensured to be sensitive to changes in both adjacent and non-adjacent spatial units. This offers a considerable enhancement in the flexibility and behavioral realism of the model specification, by allowing the testing of a whole suite of potential variables and functional form of variables affecting correlation between alternatives.

An important note is due here regarding the choice of the general formulation used to specify the allocation parameter. In this particular study, a wide variety of variables were formulated and tested to examine their impacts on the allocation parameters, as discussed in the text surrounding Equation (4). It was found that the simple distance-based measure specified in Equation (5) turned out to be statistically adequate in the context of the empirical application in this paper (that is none of the other elements of the ϕ parameter vector in Equation (5) turned out to be statistically significant). The remainder of the paper is dedicated to describing the empirical application, model estimation results, and model interpretation in the context of this

distance-based SCL (DSCL) model, which is a special case of the GSCL model proposed in the paper.

4. EMPIRICAL CONTEXT, DATA DESCRIPTION AND VARIABLE SPECIFICATION

Residential location choice is one of the most critical decisions made by a household that has a profound and lasting impact on activity-travel patterns of household members as well as non-household members with whom a household interacts. Residential location choices of households have impacts on the evolution of the built environment as transport, land use, and urban form change in response to the needs of the resident population. The study of residential location choice is of major interest at the nexus of transportation and land use modeling and is considered fundamental to the development of integrated land use – transportation modeling systems (regardless of whether the transportation model is activity-based). Guo (2004) and Bhat and Guo (2004) provide extensive literature reviews on residential location choice modeling; reviewing or citing the vast body of literature on residential location choice modeling is beyond the scope of this paper.

The residential location choice problem and the variables characterizing the choice alternatives are inherently spatial and therefore impacted by spatial effects. However, in the context of discrete choice models, there have been relatively few studies incorporating spatial effects into location choice models. The modeling methodology presented in this paper constitutes a behavioral approach to endogenously incorporate spatial dependency or spatial correlation in the model structure.

The modeling methodology is applied in this paper to the study of residential location choice for a sample of households extracted from the 2000 San Francisco Bay Area Travel Survey (BATS). Details about the survey and sampling procedures may be found elsewhere (MORPACE International, Inc., 2002). The survey data set includes detailed information about individual and household socio-demographic and employment characteristics for over 15,000 households in the Bay Area. The data set also includes detailed information on all activity-travel episodes for a two-day period for all household members.

In addition to the data set from the 2000 BATS, several secondary data sets were used to obtain spatial and built environment variables that may characterize the choice behavior of households in the region. The data sets used in this study include:

1. Land-use and demographic coverage data,
2. Zone-to-zone travel level-of-service (LOS) data,
3. The 2000 Census of population and household summary files (SF1),
4. GIS layers of highways (including interstate, toll, national, state and county highways),
5. GIS layers of local roadways (including local, neighborhood, and rural roads),
6. GIS layers of bicycle facilities, and
7. GIS layers of businesses (shopping and grocery stores, medical facilities and personal services, automotive businesses, food stores, sports and fitness centers, parks and gardens, restaurants, recreational businesses, and schools).

Among the secondary data sets indicated above, the land-use/demographic coverage data, LOS data, and the GIS layer of bicycle facilities were obtained from the Metropolitan Commission (MTC). The land-use and demographic coverage data were obtained at the traffic analysis zone (TAZ) level, and used to characterize the demographic characteristics of the households in each zone, the urban environment, and the accessibility to work or non-work related activities. The GIS layers of highways and local roadways were obtained from the 2000 Census Tiger Files. The GIS layers of businesses were obtained from the InfoUSA business directory. Further, various average household and region characteristics were obtained at the TAZ level by aggregating census block-group level data. A detailed description of these secondary data sets and the data extraction and merging procedures is provided in Bhat and Guo (2007).

The sample of households extracted for use in this study is that with residential locations in San Mateo County in the San Francisco Bay Area. In particular, the county-specific subsample consisted of 702 households which, for the purposes of this study, are assumed to have the choice of residing in one of the 115 TAZs that comprise San Mateo County.

The seven secondary data sources provided a rich set of built environment variables for each TAZ. In addition to the direct impacts of built environment variables on residential location choice, the interaction effects between household socio-demographics and built environment variables are considered to account for the taste variations of households to zonal attributes. The following categories of variables were considered for inclusion in the model specifications.

- *Zonal Land Use Structure Variables* include housing-type measures (proportion of single-family, multi-family, duplex, and other dwelling units), land use composition

measures (proportion of zonal area devoted to residential, commercial, and other land uses), and a land use diversity index (between 0 and 1 where zones with greater land use diversity obtain higher index values) computed from land use composition variables.

- *Zonal Size and Density Measures* include total population, number of housing units, population density, household density, location indicators (CBD, suburban, urban, rural), and employment density by employment type or category.
- *Regional Accessibility Measures* include measures that capture the amount of employment, shopping, and other activity opportunities that can be reached by auto and transit.
- *Zonal Ethnic Composition Measures* include variables that describe the ethnic/racial composition of the population in each zone.
- *Zonal Demographic and Housing Cost Variables* include such variables as average household size, median household income, and median housing cost in each zone.
- *Zonal Commute-related Variables* include variables that measure the commute time and cost as aggregate values across all workers in the household based on the assumption that employment location is predetermined.
- *Zonal Activity Opportunity Variables* include variables constructed from the InfoUSA business directory to describe the composition of the zones with respect to intensity or density of various types of activity opportunities including shopping, physically active recreation (e.g., gym), physically passive recreation (e.g., movie theater), natural recreation (e.g., park), and eat out.
- *Zonal Transportation Network Measures* include variables that describe highway density (lane-miles per square mile), local roadway density, bikeway density, transit availability, and street-block density.
- *Interaction Effects Variables* include those that capture interaction effects between household socio-demographics and zonal attributes such as the absolute difference between the median zonal income and household income, absolute difference between zonal average household size and household size, and household income interacted with accessibility measures.
- *Spatial Correlation Variables* include those that characterize spatial correlation patterns across alternatives and may be used to formulate alternate specifications of the allocation

parameter that represents the degree of dependency across choice alternatives, as discussed in the text following Equation (4).

5. MODEL ESTIMATION RESULTS

The final model specification was based on intuitive considerations, insights from previous literature, parsimony in specification, and statistical fit/significance considerations. In order to empirically demonstrate the value of incorporating spatial correlation across contiguous and non-contiguous zones, three different residential location choice models were estimated. These include:

1. A standard MNL model in which the residential choice alternatives are considered uncorrelated
2. A SCL model in which the residential choice alternatives are allowed to be endogenously correlated only across adjacent or contiguous zones
3. A GSCL model in which the residential choice alternatives are allowed to be endogenously correlated across contiguous and non-contiguous zones

As noted earlier in the paper, the MNL model is a restricted version of the SCL and GSCL models. Therefore, to allow a behaviorally and statistically sound comparison of models, each model is estimated with the same sample data and set of variables. Table 2 presents model estimation results for the MNL and GSCL models. It should be noted that, in this particular application context, the dissimilarity (or correlation) parameter in the SCL model is found to be not statistically significantly different from one. When the dissimilarity parameter is equal to one, the SCL model reduces to the standard MNL model. Therefore, the MNL and SCL model estimation results are one and the same in this paper. Also, the GSCL model specification collapsed to the DSCL model, and so we will henceforth substitute the acronym DSCL for the GSCL model (including in Table 2).

In general, from a qualitative examination of the model estimation results in Table 2, the MNL and DSCL models offer similar behavioral interpretations. Coefficient values and signs are generally consistent with respect to behavioral interpretation across the two model specifications. In both models, it is found that households do not tend to locate or reside in zones with greater degrees of land use mix. This is further corroborated by the finding that households tend to be positively inclined to locate in zones with larger numbers of households. These

findings are consistent with the pattern of land use development where households tend to reside in homogeneous residential zones or subdivisions with little mix of other land uses. A clustering effect is observed for the Caucasian population, as evidenced by the positive coefficient associated with the interaction variable capturing the tendency of Caucasian households to locate themselves in predominantly Caucasian zones. Such a clustering effect is not found for other ethnic groups in this particular sample. Additional clustering effects are observed with respect to zonal demographics and housing cost. Households tend to locate in zones with similar income levels and household sizes. As the difference between the household income or household size and that of the zone increases, the likelihood that the household will reside in that zone decreases. Housing affordability is another major factor influencing household residential location choice. Model estimation results suggest that, as the median housing value rises, the likelihood of that zone being chosen by a household as a residential location falls. The effect of the zonal activity opportunity measures reflects the positive influence of the physically active recreation centers such as fitness centers, sports centers, dance, and yoga studios. As the number of these physical activity centers increases in a zone, households are more likely to reside in such a zone. Consistent with several other findings, residential location choice is negatively associated with highway density. Once again, considering that the general pattern of residential development/location is one where households reside in low density suburban locations, this finding of a negative association with highway density is consistent with expectations. Commute-related variables are important factors in household residential location choice. As the total commute time (aggregated over all commuters in a household) rises, the likelihood of residential location in a particular zone falls. On the other hand, residential location choice is positively impacted by the availability of transit between the home and work zones. As the number of commuters in the household who have transit connectivity increases, the likelihood of residential location in a zone increases as well. This suggests that transit availability can play a role in shaping residential location choices of households. The combination of the effects of these two commute-related variables, coupled with the high and rising fuel prices seen in the last few years, makes the case for enhancing transit service availability, enhancing residential location opportunities along transit lines, and mixing land uses so that commute times are reduced. Such measures may result in households locating in neighborhoods that foster the use of alternative modes of transportation and reduce the amount of distance and time involved in

commuting. It is interesting to note that, in this particular application context, regional accessibility measures were not found to be statistically significant in explaining residential location choice. As the transportation system is generally of a ubiquitous nature in virtually all urban areas in the United States, it is not surprising that these variables have not turned out statistically significant in explaining residential location choice.

Two key parameters of interest in the context of this study are the dissimilarity or correlation parameter, μ , and the distance coefficient in the allocation parameter, ϕ . The dissimilarity or correlation parameter for the MNL model is 1 because the MNL model assumes independence across all choice alternatives. In this particular context, it was found that the SCL model offered the same result. The dissimilarity parameter for the DSCL is 0.617 and this parameter is statistically significantly different from 1 at the 0.05 level of significance. The finding of a dissimilarity parameter that is statistically significantly smaller than one indicates that the DSCL model rejects the MNL (and SCL) models, and that there is a high level of spatial correlation across contiguous and non-contiguous zones in residential location choice, which both the MNL and SCL fail to recognize in this particular application. In addition, the distance coefficient in the allocation parameter, ϕ , is also found to be statistically significant at the 0.05 level of significance. The parameter is negative suggesting that the degree of correlation decreases as the distance between zones increases. It is also interesting to note that this coefficient is approximately equal to a value of 2, which is similar to a gravity model-type formulation where the decay function assumes that the interaction between two zones falls off as the square of the distance between them. This parameter does not exist in the MNL and SCL models; given that it is found to be statistically significant in this study, the DSCL is capturing a distance-based correlation effect between zones that is completely ignored in the MNL and SCL models. The DSCL model is also found to offer a statistically superior goodness-of-fit in comparison to the MNL/SCL models.

Differences between the MNL/SCL and DSCL model are further elucidated through a comparison of elasticity effects or measures offered by the alternative model specifications. Tables 3 and 4 present direct and cross-elasticity effects estimated for a randomly selected household in the survey subsample. The elasticities are computed for the randomly chosen household with respect to changes in attributes of the residential location (zone) where the household currently resides. Thus, the direct and cross elasticities reflect the change in

likelihood of choosing a zone as a residential location in response to change in attribute of the zone where the household currently resides. Cross elasticities are computed for four alternate zones where two zones are contiguous or adjacent to the chosen residential zone and two zones are non-contiguous. The four zones are considered in the tables in ascending order of distance from the chosen residential zone.

Direct elasticity values are provided in the first column of the tables. Just as the model coefficients in Table 2 showed differences between the MNL/SCL and DSCL models, the direct elasticity values also show some differences. As the DSCL model incorporates a different spatial correlation structure, these differences are consistent with expectations. Although the magnitudes and signs of direct elasticities are generally consistent across the model specifications, one should recognize that these elasticity differences could have substantial consequences when applied to a large population of households.

More significant and noteworthy in the context of this study is the comparison of cross-elasticity values. In Table 3, it should be noted that there is only a single cross-elasticity value across all four alternate zones because the SCL model reduced to the MNL model in this particular specification. If the SCL model offered a dissimilarity parameter significantly less than one, then the SCL model would have offered two sets of cross-elasticity values – one set for the two contiguous zones and one set for the two non-contiguous zones. On the other hand, the DSCL model presents four distinct cross-elasticity values because the degree of spatial correlation between the chosen subject zone and the four alternate zones is dependent on the distance between the zones. Cross-elasticity values would be identical only in instances where the spatial separation between zones under consideration was equal.

An examination of the cross-elasticity measures in Table 4 shows that the cross-elasticity decreases as the distance between the chosen subject zone and the alternate zone increases. This continuous variation in cross-elasticity is a manifestation of the distance-based specification of the spatial correlation allocation parameter in the DSCL model. The DSCL model is a more flexible version of the SCL model in that it accommodates disproportionate shifts in residential location choice probabilities in response to changes in built environment and zonal attributes. These findings suggest that the DSCL is a behaviorally realistic and statistically robust representation of spatial correlation across choice alternatives in location choice modeling contexts.

6. CONCLUSIONS

Location choice behavior of households and individuals lies at the heart of activity-travel demand modeling. Activity-travel behavior is characterized by interactions in time and space and it is widely recognized that the incorporation of time-space interaction effects is critical to understanding, explaining, and modeling activity-travel demand under a wide range of land use, transport, technology, and policy scenarios. Although great strides have been made in recognizing temporal constraints and dimensions in activity-travel demand modeling, the incorporation of spatial effects in modeling frameworks has generally lagged advances made in the temporal domain. This paper contributes to the profession's ability to reflect spatial effects in discrete choice models of travel behavior.

This paper focuses on the issue of spatial correlation that pervades most location choice modeling contexts where choice alternatives are correlated with one another. In many instances, location choice alternatives are correlated with one another due to unobserved spatial and demographic attributes that lead to spatial dependency or correlation. In the presence of correlation across choice alternatives over space, the assumption of independency of choice alternatives intrinsic to the classic multinomial logit (MNL) model is violated and parameter estimates of standard logit models will be biased and inconsistent. Although there are several methods to try and accommodate spatial correlation, many of these methods have limitations and require the analyst to have a priori knowledge about the nature of the spatial correlation prevalent in the choice set. Existing methods are also unsuitable to deal with contexts where the location choice set is very large, a situation that is often encountered in travel demand modeling.

In previous work, Bhat and Guo (2004) proposed a spatially correlated logit (SCL) model that accommodated correlations across location alternatives that were contiguous to one another. As long as two spatial units (such as traffic analysis zones) shared even a tiny stretch of common boundary, the alternatives were considered correlated. If two alternatives did not share a common boundary (*i.e.*, they were non-contiguous), they were treated as being completely uncorrelated. Although this approach accommodated spatial correlation between neighboring zones, it did not accommodate spatial correlation between non-contiguous zones and was sensitive to the configuration of zonal boundaries. If the zonal configuration was modified, the spatial correlation structure in the SCL model could change drastically leading to unreasonable degrees of change in parameter estimates.

In order to overcome these limitations, this paper offers a generalized version of the SCL model (labeled the Generalized SCL or GSCL model) by considering spatial correlation across all zone pairs. A mixed version of the GSCL (MGSCl) model is also formulated and presented in this paper to demonstrate how the GSCL model can be enhanced to accommodate random taste variations across decision-makers.

The GSCL model is estimated and compared to the standard MNL and SCL in the context of a residential location modeling application. A sample of about 700 households in the San Mateo County of the San Francisco Bay Area is extracted from the 2000 San Francisco Bay Area Travel Survey (BATS). Extensive set of secondary built environment variables were appended to the survey records to form a comprehensive database for model estimation. The GSCL model takes a simpler distance-based SCL (or DSCL) model formulation in the empirical estimation, and clearly rejects the standard MNL specification. On the other hand, in the particular empirical context considered in this study, the SCL offered a dissimilarity parameter not significantly different from one, making it essentially equivalent to the standard MNL. The distance coefficient in the spatial correlation allocation parameter in the DSCL model is approximately -2, suggesting that the degree of correlation between location alternatives in the residential choice context falls as the second power of the distance between them. Comparisons of cross-elasticities further demonstrate that the DSCL is able to capture the greater interaction effects between zones that are closer together, while simultaneously recognizing that zones that are farther apart (and non-contiguous) are also correlated, albeit to a lesser degree.

In conclusion, the GSCL model offers a rigorous approach for incorporating a continuous spatial correlation structure in discrete choice models of location choice. Activity-based travel models that purport to capture time-space interactions should be formulated to recognize these spatial correlation effects. Ongoing research efforts include extending the GSCL model to joint discrete-continuous simultaneous equations modeling contexts, and testing the robustness of the findings in other empirical contexts.

ACKNOWLEDGMENTS

This research was funded in part by Environmental Protection Agency Grant R831837. The authors are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

- Abbe, E., M. Bierlaire, and T. Toledo (2007) Normalization and Correlation of Cross-Nested Logit Models. *Transportation Research Part B*, 41(7), 795-808.
- Abraham, J.E., and J.D. Hunt (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.
- Bekhor, S, and J.N. Prashker (2008) GEV-based Destination Choice Models that Account for Unobserved Similarities among Alternatives. *Transportation Research Part B*, 42(3), 243-262.
- Bhat, C.R., and J.Y. Guo (2004) A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. *Transportation Research Part B*, 38(2), 147-168.
- Bhat, C.R., and J.Y. Guo (2007) A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. *Transportation Research Part B*, 41(5), 506-526.
- Bhat, C.R., and I.N. Sener (2008) A Copula-Based Closed-Form Binary Logit Choice Model for Accommodating Spatial Correlation Across Observational Units. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, June.
- Bhat, C.R., and H. Zhao (2002) The Spatial Analysis of Activity Stop Generation. *Transportation Research Part B*, 36, 557-575.
- Bolduc, D., B. Fortin, and M. Fournier (1996) The Effect of Incentive Policies on the Practice Location of Doctors: A Multinomial Probit Analysis. *Journal of Labor Economics*, 14(4), 703-732.
- Boots, B.N., and P.S. Kanaroglou (1988) Incorporating the Effect of Spatial Structure in Discrete Choices Models of Migration. *Journal of Regional Science*, 28, 495-507.
- Daly, A., and M. Bierlaire, (2006) A General and Operational Representation of Generalized Extreme Value Models. *Transportation Research Part B*, 40, 285-305.
- Daly, A.J., and S. Zachary (1978) Improved multiple choice models. In *Determinants of Travel Choice*, edited by D.A. Hensher and M.Q. Dalvi, Saxon House, Westmead.
- Deng, Y., S.L. Ross, and S.M. Wachter (2003) Racial Differences in Homeownership: The Effect of Residential Location. *Regional Science and Urban Economics*, 33, 517-556.
- Fleming, M. (2004) Techniques for Estimating Spatially Dependent Discrete Choice Models. In *Advances in Spatial Econometrics*, Eds. Raymond Florax and Luc Anselin, Springer, Berlin.
- Fotheringham, A.S., and C. Brunson (1999) Local Forms of Spatial Analysis. *Geographical Analysis*, 31(4), 340-358.
- Garrido, R.A., and H.S. Mahmassani (2000) Forecasting Freight Transportation Demand with the Space-time Multinomial Probit Model. *Transportation Research Part B*, 34(5), 403-418.
- Guo, J.Y. (2004) Addressing Spatial Complexities in Residential Location Choice Models. Doctoral Dissertation, Department of Civil Engineering, The University of Texas at Austin, Texas.
- Haynes, K., and A.S. Fotheringham (1990) The Impact of Space on the Application of Discrete Choice Models. *Review of Regional Studies*, 2, 39-49.

- Horowitz, J.H. (1981) Identification and Diagnosis of Specification Error in the Multinomial Logit Model. *Transportation Research Part B*, 15, 345-360.
- Hunt, L.M., B. Boots, and P.S. Kanaroglou (2004) Spatial Choice Modelling: New Opportunities to Incorporate Space into Substitution Patterns. *Progress in Human Geography*, 28(6), 746-766.
- Koppelman, F.S., and V. Sethi (2008) Closed-form Discrete Choice Models. *Handbook of Transport Modelling, 2nd edition*, Ch. 13, 211-225, edited by D.A. Hensher and K.J. Button, Elsevier Science.
- Marzano, V., and A. Papola (2008) On the Covariance Structure of the Cross-Nested Logit Model. *Transportation Research Part B*, 42(2), 83-98.
- McFadden, D. (1978) Modeling the Choice of Residential Location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds) *Spatial Interaction Theory and Planning Models*, North Holland, Amsterdam, The Netherlands, pp. 75-96
- McFadden, D. (1981) Econometric Models of Probabilistic Choice. In Manski, C. and McFadden, D., (Eds), *Structural Analysis of Discrete Data*, MIT Press, Cambridge, MA, pp. 198-272.
- Miyamoto, K., V. Vichiensan, N. Shimomura, and A. Páez (2004) Discrete Choice Model with Structuralized Spatial Effects for Location Analysis. Presented at the 83rd Annual Conference of the Transportation Research Board, Washington, D.C.
- MORPACE International, Inc. (2002) Bay Area Travel Survey Final Report. Metropolitan Transportation Commission, CA.
<ftp://ftp.abag.ca.gov/pub/mtc/planning/BATS/BATS2000/>
- Openshaw, S. (1984) *Concepts and Techniques in Modern Geography: Number 38 - The Modifiable Areal Unit Problem*, Geo Books, Norwick.
- Páez A., D.M. Scott, and E. Volz (2008) A Discrete-Choice Approach to Modeling Social Influence on Individual Decision Making. *Environment and Planning B: Planning and Design*, 35(6), 1055-1069.
- Pellegrini, P.A., and A.S. Fotheringham (2002) Modeling Spatial Choice: A Review and Synthesis in a Migration Context. *Progress in Human Geography*, 26(4), 487-510.
- Tobler, W. (1970) A Computer Model Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(2), 234-240.
- Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press.
- 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. <http://www.odot.state.or.us/tddtpan/modeling.html>
- Wen, C-H., and F.S. Koppelman (2001) The Generalized Nested Logit Model. *Transportation Research Part 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 A*, 9(3), 285-344.

LIST OF TABLES

Table 1. Expressions for the Direct and Cross-Elasticities in the MNL, SCL, and GSCL Models

Table 2. Estimation Results for the MNL and DSCL Models

Table 3. Disaggregate Elasticity Effects for the MNL/SCL Model

Table 4. Disaggregate Elasticity Effects for the GSCL Model

Table 1. Expressions for the Direct and Cross-Elasticities in the MNL, SCL, and GSCL Models

Model	Direct elasticity [§]	Cross-elasticity ^{**}
MNL	$(1 - P_i)\beta_m x_{im}$	$- P_i \beta_m x_{im}$
SCL	$\left\{ \sum_{j \neq i} P_{ij} P_{ij} \left[(1 - P_i) + \left(\frac{1}{\mu} - 1 \right) (1 - P_{ij}) \right] \right\} \frac{\beta_m x_{im}}{P_i}$	$- \left[P_i + \frac{\left(\frac{1}{\mu} - 1 \right) P_{ij} P_{ij} P_{j ij}}{P_j} \right] \beta_m x_{im} \text{ if } i \text{ and } j \text{ are contiguous}$ $- P_i \beta_m x_{im} \text{ if } i \text{ and } j \text{ are not contiguous}$
GSCL	$\left\{ \sum_{j \neq i} P_{ij} P_{ij} \left[(1 - P_i) + \left(\frac{1}{\mu} - 1 \right) (1 - P_{ij}) \right] \right\} \frac{\beta_m x_{im}}{P_i}$	$- \left[P_i + \frac{\left(\frac{1}{\mu} - 1 \right) P_{ij} P_{ij} P_{j ij}}{P_j} \right] \beta_m x_{im} \text{ if } i \text{ and } j \text{ are contiguous or not contiguous}$

[§] Direct elasticity refers to the percentage change in the choice probability of alternative i due to a 1% change in the m^{th} variable associated with alternative i . Note that we are suppressing the index for the individual in this table.

^{**} Cross-elasticity is the percentage change in the choice probability of alternative j due to a 1% change in the m^{th} variable associated with alternative i .

Table 2. Estimation Results for the MNL and DSCL Models

Variables	Multinomial Logit Model (MNL) ^{††}		Distance-based Spatially Correlated Logit Model (DSCL)	
	Parameter	t-stat	Parameter	t-stat
Zonal land-use structure				
Land-use mix	-0.447	-2.24	-0.329	-1.91
Zonal size and density				
Log # households in zone	0.916	9.45	0.750	6.35
Zonal ethnic composition measure				
Fraction of Caucasian population interacted with Caucasian dummy variable	2.043	6.26	1.772	5.29
Zonal demographics and housing cost				
Absolute difference between zonal median income and household income (\$ x 10 ⁻³)	-0.015	-6.94	-0.012	-5.62
Absolute difference between zonal average household size and household size	-0.210	-2.08	-0.182	-2.05
Average of median housing value	-0.153	-5.00	-0.125	-4.26
Zonal activity opportunity variables				
Number of physically active recreation centers such as fitness centers, sports centers, dance and yoga studios	0.020	2.59	0.016	2.47
Zonal transportation network measures				
Highway density (mileage per square mile)	-0.092	-1.93	-0.054	-1.36
Household level commute variables				
Total commute time (by auto) of all commuters in the household (min.)	-0.042	-13.53	-0.040	-10.98
Number of commuters in household with home and work zones served by transit within 30 min	0.605	5.34	0.510	4.61
Dissimilarity (correlation) parameter (μ)	1.00	---	0.617	3.79
Distance coefficient in allocation parameter (ϕ)	---	---	-2.052	-2.02
Number of Observations	702		702	
Log-likelihood at convergence	-3054.06		-3051.17	

^{††} In this empirical context, the SCL model is equivalent to the MNL model.

Table 3. Disaggregate Elasticity Effects for the MNL/SCL Model

		<i>MNL and SCL Models</i>			
		Zone 1	Zone 2	Zone 3	Zone 4
Contiguity		contiguous	contiguous	non-contiguous	non-contiguous
Distance between zones (in miles)		0.86	1.05	2.42	5.58
Variables	Direct Elasticity	Cross Elasticity with respect to different zones			
Zonal land-use structure					
Land-use mix	-0.1669	0.0124			
Zonal size and density					
Log # households in zone	6.9938	-0.5199			
Zonal ethnic composition measure					
Fraction of Caucasian population interacted with Caucasian dummy variable	1.3699	-0.1018			
Zonal demographics and housing cost					
Absolute difference between zonal median income and household income (\$ x 10 ⁻³)	-0.2080	0.0155			
Absolute difference between zonal average household size and household size	-0.2669	0.0198			
Average of median housing value	-0.8136	0.0605			
Zonal activity opportunity variables					
Number of physically active recreation centers such as fitness centers, sports centers, dance and yoga studios	0.2737	-0.0203			
Zonal transportation network measures					
Highway density (mileage per square mile)	-0.0885	0.0066			
Household level commute variables					
Total commute time (by auto) of all commuters in the household (min.)	-0.1082	0.0080			
Number of commuters in the household with home and work zones served by transit within 30 min	0.5630	-0.0418			

Table 4. Disaggregate Elasticity Effects for the GSCL Model

		<i>DSCL Models</i>			
		Zone 1	Zone 2	Zone 3	Zone 4
Contiguity		contiguous	contiguous	non-contiguous	non-contiguous
Distance between zones (in miles)		0.86	1.05	2.42	5.58
Variables	Direct Elasticity	Cross Elasticity with respect to different zones			
Zonal land-use structure					
Land-use mix	-0.1306	0.0136	0.0129	0.0097	0.0088
Zonal size and density					
Log # households in zone	6.0902	-0.6351	-0.5998	-0.4534	-0.4102
Zonal ethnic composition measure					
Fraction of Caucasian population interacted with Caucasian dummy variable	1.2634	-0.1318	-0.1244	-0.0941	-0.0851
Zonal demographics and housing cost					
Absolute difference between zonal median income and household income ($\$ \times 10^{-3}$)	-0.1866	0.0195	0.0184	0.0139	0.0126
Absolute difference between zonal average household size and household size	-0.2462	0.0257	0.0242	0.0183	0.0166
Average of median housing value	-0.7100	0.0740	0.0699	0.0529	0.0478
Zonal activity opportunity variables					
Number of physically active recreation centers such as fitness centers, sports centers, dance and yoga studios	0.2346	-0.0245	-0.0231	-0.0175	-0.0158
Zonal transportation network measures					
Highway density (mileage per square mile)	-0.0546	0.0057	0.0054	0.0041	0.0037
Household level commute variables					
Total commute time (by auto) of all commuters in the household (min.)	-0.1111	0.0116	0.0109	0.0083	0.0075
Number of commuters in the household with home and work zones served by transit within 30 min	0.5044	-0.0526	-0.0497	-0.0376	-0.0340