# Modeling Spatial Discrete Choice

Oleg A. Smirnov
Department of Economics
University of Toledo,
2801 W Bancroft St
Toledo, OH 43606-3390
*email: oleg.smirnov@utoledo.edu*

**Abstract**

The paper presents a basic spatial discrete choice modeling framework obtained by applying random utility theory to discrete choices made by heterogeneous spatially dependent individuals. The newly developed framework has two main advantages over existing approaches. First, individual decision-makers are no longer assumed to be independent and non-interacting but spatially interdependent in their preferences facilitating the development of applied discrete choice models using a wide range of spatial data. Second, pseudo maximum likelihood estimator is developed for this model that is consistent and computationally feasible for large datasets. The performance of the pseudo maximum likelihood estimator for the spatial discrete choice model is illustrated using simulated data.

*Key words:* spatial interdependence, discrete choice, maximum likelihood, spatial random utility

*JEL: C2, C5, R0*

# 1 Introduction

Discrete choice models have gained popularity as useful tools for studying transportation choice modeling (McFadden 1978; Koppelman and Sethi 2005), residential choice location (Bhat and Guo 2004; Bekhor and Prashker 2008), social interactions (Brock and Durlauf 2001), marketing (Chakravati et al 2005), resource management and recreational choices (Haffen, Massey, and Adamowicz 2005), other applications (Train 2003; Louviere, Train, Ben-Akiva, Bhat et al 2005 with references), theoretical analysis of discrete choice games (Aguirregabiria and Mira 2007), and discrete choice experiments (Louviere, 2006) among others. The unifying feature of discrete choice models is the focus on the behavioral aspects of individual decision-making and the assumption that individuals are independent. The application of discrete choice models to spatial data, however, necessitates the development of discrete choice frameworks that would allow for spatial effects. The main purpose of this article is to develop a spatial random utility framework including a consistent estimation method that is not based on simulation.

Out of the two major groups of spatial effects – spatial dependence and spatial heterogeneity – that are encountered in spatial data, heterogeneity of individual observations generally does not contradict the assumption of independence. Since the heterogeneity of individuals entails no new behavioral concepts or phenomena, both heterogeneity of preferences and heterogeneity between alternatives perfectly fit the random utility framework. In contrast, spatial dependence entails interdependence between individuals, interdependence between their preferences, and therefore, choices. Arguably, spatial dependence is a particularly important spatial effect in spatial discrete choice models, as spatial interdependence of individuals presents a new aspect of individual behavior not found in the models with non-interactive independent individuals. Spatial interdependencies between individuals affect their preferences, creating the phenomenon of socially influenced decision-making, so that individuals neither act fully independently, nor reach decisions jointly. Studying the effects of spatial dependence on discrete choices is important for extending discrete choice modeling to the analysis of the social aspects of individual decision-making. This paper focuses on spatial dependence in discrete choice models.

Earlier attempts to develop and apply discrete choice models to spatial data have been overshadowed by a lack of common theoretical foundations to modeling spatial discrete choice, confusion about how to treat spatial effects in such models, and insurmountable computational difficulties of calibrating and analyzing the results of some models (Conley and Topa 2007; Graham 2008; Mohammadian, Haider and Kanaraglou, 2005; Klier and McMillen 2008; Murdoch, Sandler and Vijverberg 2003; Smith and LeSage 2004; among others). The major complication to an easy development of spatial discrete choice models is the fact that most spatial interdependencies are not directly observed by the researcher. In addition, most decision-makers act with partial information, prompting economists to assume that individuals form probabilistic expectations and maximize utility (Manski 2004). As it concerns spatially dispersed individuals, the researchers hypothesize that spatial interdependencies are a form of adaptation that individuals discretionally deploy in order to maximize their utility. The exact mechanism by which spatial dependence thrives in observed choice behavior is application-specific and varies along with the definition of the decision-makers (households, firms, or other entities), and is distinctly related to the essence of both cognitive and rational aspects of individual decision-makers. Casting spatial dependence in models and measuring spatial effects in practice are non-trivial tasks that require appropriate methodological techniques and prudence in interpreting results. To date, the maximum likelihood estimator for spatial discrete choice models is not computationally feasible, and commonly used calibration procedures are based on simulations or approximations.

To focus on the social aspects of the behavior of individual decision-makers, spatial interdependencies are modeled at the level of individual preferences to reflect interdependencies of individuals prior to making choices. Spatial interdependencies are not directly observed by the researcher. Instead, the researcher observes individual discrete choices, which are used to measure the effect of spatial interdependencies. The paper is structured as follows. Section 2 provides a brief overview of recent developments in modeling spatial effects in discrete choice models. Section 3 introduces the basic spatial discrete choice model. Section 4 develops a pseudo maximum likelihood estimator for the spatial discrete choice model and relevant computational issues. In Section 5, simulated data are used to illustrate the performance of the estimator for various dataset sizes and values of spatial autoregressive coefficient. The conclusion summarizes the findings.

# 2 Spatial Effects in Discrete Choice Models: Recent Developments

It is possible to group recent developments in the literature on spatial effects in discrete choice models by the type of spatial effect and the manner in which it is incorporated in the model.

*Spatial dependence between alternatives.* Bekhor and Prashker (2008), Bhat and Guo (2004), Koppelman and Sethi (2005), and Sivakumar and Bhat (2007) among others, present specific cases of the random utility models (McFadden 1978; Train 2003; Ben-Akiva and Lerman 1985). These models allow for some spatial dependence between alternatives, which is commonly resolved by nested logit specifications. The constant elasticity of substitution in the error terms is structured to represent the substitution effects in the individual preferences between the alternatives according to their mutual spatial arrangement.

*Spatial dependence in the linear probability model.* The model of Pinkse, Slade, and Shen (2006) represents an attempt to circumvent modeling the decision rule and individual preferences. The binary choice decision variable is given as the sum of discrete and continuous variables as opposed to the decision rule in discrete choice variables. Its advantage is in the simplicity of a linear probability single-equation model, which can be estimated by GMM (Pinkse, Slade, and Shen 2006). However, the oversimplified linear probability model confuses discrete choice variables with conditional choice probabilities, effectively restricting model use and analysis. Additional pitfalls include apparent asymptotic biases of both GMM and GEL estimators for this type of model (Iglesias and Phillips 2008).

*Spatial probit and spatial logit.* Anselin (2002) mentioned distinctive ways to model spatial dependence in discrete choice models. Comparing probit and logit frameworks, he noted that the spatial probit has an advantage over the spatial logit because the error term is analytically intractable in the latter. It is worth noting, however, that probit is a binary choice model that can not be easily extended to the cases of choice sets with more than two alternatives, and some estimation methods such as GMM are not sensitive to the exact specification of the error term.

The spatial probit specification is more popular in the literature than the spatial logit. Various moment-based and likelihood-based techniques proposed and tested in various settings. McMillen (1992) pioneered the application of the expectation-maximization (EM) algorithm for estimating the coefficients – albeit not the information matrix – of the spatial probit model. Pinkse and Slade (1998) examined conditions for applying GMM to the spatial probit model, analyzing asymptotic properties of the GMM estimator for a subset of model specifications. The pivotal requirement for the method is the existence of appropriate instruments. Murdoch, Sandler and Vijverberg (2003) developed the spatial probit model and used it as the first stage to describe a two-stage game, in which the decision-makers commit to discrete choice in the first stage and decide on the continuous variable in the second. A key element of the model is the spatial dependence in the latent variables, while the error terms are essentially independent. Spatial probit with spatial dependencies embedded in the error term also have been developed by Smith and LeSage (2004) in the context of Bayesian inference. In their model, the compound error terms consist of non-spatial and spatial components. Several techniques for dealing with spatial dependence in the spatial probit model are reviewed in Fleming (2005).

The spatial logit specification of Klier and McMillen (2008) explicitly assumes that choice probability is given by the closed-form expression as typically done in non-spatial multinomial logit models, notwithstanding spatial dependencies in the error terms. They show that this clever shortcut allows an easy extension of Pinkse and Slade's (1998) estimator to the multinomial logit model. Finding GMM to be impractical for large datasets, Klier and McMillen (2008) propose a few approximations of sample-based moments computations. The approach produces a practical estimation method without the need to second-guess on the choice of instrumental variables. The drawback of this approach is that the asymptotic properties of the GMM are no longer applicable.

*Statistical mechanics approach.* Brock and Durlauf (2001, 2002, 2007) imported a model of interactions from statistical mechanics and applied it to the analysis of social and spatial interactions. The focus of their approach is on the relation between individual decisions and neighbors' characteristics and expectations of neighbors' decisions. The model analysis focuses on finding the equilibrium for these expectations. Since expectations are non-random, the interactions between random components of individuals preferences are left out of the scope of the model. These simplifications allow the characterization of the equilibrium state for the error term drawn from the extreme value distribution (Brock and Durlauf 2002). Despite analytical advancements, the apparent absence of empirical studies utilizing model conforms with Anselin's (2002) suggestion that spatial probit is easier to adapt to practical use.

*Autologistic model.* As an alternative to the random utility model (Ben-Akiva and Lerman 1985; McFadden 1978; Train 2003), spatial dependencies between discrete variables can be modeled without using the notion of preferences or individual utilities. This approach was popularized by Besag (1974) with the specification of the autologistic model. In the model, the probability of discrete variable is conditional on the characteristics of individual or location and spatially weighted discrete variable. This model neither assumes nor requires the notion of preferences, and hence, any volitional action such as choice – the discrete variable is given by the probability distribution. For this reason it is very popular in ecological modeling and other non-economic applications (Dormann 2007; Ward and Gleditsch 2002).

*Variations of autologistic model.* Examples of such settings include Mohammadian, Haider and Kanaroglou (2005), Páez and Scott (2007), Páez, Scott, and Volz (2008). A common feature of these models is the placement of neighbors' choices in the set of explanatory variables affecting individual utility; i.e., neighbor's choices are known to each individual decision-maker. The probability of discrete choice is conditional on individual-specific variables and neighbors' choices rather than utility. In this case, preferences of neighbors are unknown to individuals, interdependencies are purely probabilistic, and the model is analytically equivalent to the autologistic model; i.e., the decision rule and the expression for the individual utility are folded into the autologistic model. The simultaneity of choices makes these models appear self-contradictory, as all neighboring individuals' choice are conditioned on each other.

*Endogenous spatial weights.* The analysis of local area unemployment by Conley and Topa (2007) uses the notion of switching behavior, so that only the unemployed rely on spatial interactions while seeking employment. Thus, the notion of neighborhood varies with the employment status of the agents: for the employed, the neighborhood is irrelevant, while for the unemployed, the probability of finding a job is affected by the employment status of social neighbors. Individual characteristics of individuals are relevant uniformly. One might view this model as an extension of the autologistic model in which spatial interdependencies are endogenous. Specifically, the neighbors' discrete choices have varying effects depending on the individual's employment status. Since the spatial weights matrix is determined by the dependent discrete variable and thus is endogenous, Conley and Topa (2007) use a simulation-based calibration method.

*Spatial heterogeneity.* Spatial heterogeneity in discrete choice models is routinely modeled by spatial panel data models with either fixed or random effects for grouped data. In these models (Case 1992; Dugundji and Walker 2005; Dugundji and Gulyás 2008; Walker and Li 2007) individuals are pooled into groups based on their location and it is assumed that their interactions are limited to within-the-group interactions. This leads to group-wise heterogeneity. While the study of group-effects might be meaningful in the context of social interactions, where social group membership effectively channels interactions between individuals to within the group interactions, this approach is unproductive for the study of spatial dependence.

# 3 Basic Spatial Discrete Choice Model

The entire society is given by the finite set of individuals $\mathcal{N}$, $|\mathcal{N}| = n$. Each individual chooses one and only one alternative from the set $\mathcal{M}$, $|\mathcal{M}| = m$. Note that $\mathcal{M} \cap \mathcal{N} = \emptyset$. Spatial dependence between individuals in this paper is understood as the spatial dependence between individual preferences. In all other respects decision-makers are independent and their choices have no social effects. Individual preferences are modeled by utility, which is assumed to be additive. Let $u_{qj}$ be random utility of agent $q$ from selecting alternative $j$. Let $u_j = (u_{1j}, u_{2j}, \ldots, u_{nj})'$ be the $n \times 1$ vector of agents' random utilities from alternative $j$, $v_j = (v_{1j}, v_{2j}, \ldots, v_{nj})'$ the vector of private deterministic components of the individual utilities, and $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{nj})'$ the vector of private stochastic components of individual utilities from alternative $j$.

The basic spatial random utility model is

$$u_j = \rho W u_j + v_j(\beta) + \varepsilon_j, j \in \mathcal{M}, \tag{1}$$

where $\rho \in \Theta_\rho$ and $\beta \in \Theta_\beta$ are model parameters. In this model, only linear effects of interactions are captured. The interactions are defined by the $n \times n$ non-negative spatial weights matrix $W$. For each individual $j \in \mathcal{N}$, the set of non-zero spatial weights $\mathcal{N}_j \subset \mathcal{N}$ defines spatial neighborhood for $j$: $k \in \mathcal{N}_j \Leftrightarrow W_{jk} > 0$. A zero entry in the matrix $W$ indicates a non-interacting pair of individuals, and a positive $W_{tr} > 0$ indicates that individual $r$ affects the utility of individual $t$. The structure of non-zero weights is symmetric: $W_{ij} \neq 0 \Leftrightarrow W_{ji} \neq 0$ and anti-reflexive: $W_{ii} = 0, i \in \mathcal{N}$. The specific definition of neighborhood might vary by application, as it is important to select the definition of neighborhood that corresponds to the nature of spatial interactions. Common approaches to constructing spatial weights based on geographic features include contiguity and distance-based techniques (Anselin 2006). For simplicity of further analysis, the spatial weights matrix $W$ is row-standardized so that its largest eigenvalue equals 1.

With the row-standardized spatial weights matrix, the necessary condition for the model (1) to represent stationary spatial process, is that the coefficient $\rho$ is bounded by the parameter domain $\Theta_\rho = (1/\omega; 1)$, where $\omega$ is the lowest eigenvalue of the spatial weights matrix $W$. This condition ensures that the transformation matrix $I - \rho W$ is non-singular, and its inverse $(I - \rho W)^{-1}$ is a positive definite matrix. The parameter domain $\Theta_\rho$ allows for positive, negative, and zero values of the autoregressive coefficient $\rho$. If $\rho > 0$, interdependencies between individuals can be characterized as cooperative, individual preferences exhibit substantial internal similarities, partial effect of neighbors' utility on each individual is positive value $\rho W_{tr}$. Negative $\rho$ indicates individual preferences are repulsive, as if individuals aim to underscore their individuality within the neighborhood by not following preferences dominant in the neighborhood. A zero value of $\rho$ suggests lack of substantial interdependence, or independence of individual preferences.

The decision rule is consistent with the notion of rationality of individual decision-makers. Denote $y_j = (y_{1j}, y_{2j}, \ldots, y_{nj})'$ the $n \times 1$ vector of discrete choices with regard to alternative $j \in \mathcal{M}$. The relation between individual utilities and chosen alternatives is

$$y_{qj} = \begin{cases} 1, \text{ if } u_{qj} \geq u_{qi}, i \in \mathcal{M} \\ 0, \text{ otherwise} \end{cases}, \tag{2}$$

so that one and only one alternative is chosen by each individual: $\sum_j y_{qj} = 1$, $\forall q \in \mathcal{N}$.

Finally, the stochastic component is important as it specifies the context in which the individual makes decisions. Suppose an individual has numerous opportunities (combinations of unobserved circumstances) to make a choice. Each opportunity $k$ is associated with utility $\varepsilon_{qjk}$. Opportunistic utilities $\varepsilon_{qjk}$ are assumed to be i.i.d. Opportunities available to one individual are unrelated to those available to others. A rational individual chooses utility-maximizing opportunity $m : \varepsilon_{qjm} = \max_k\{\varepsilon_{qjk}\}$. Denote statistical distribution of $\varepsilon_{qjm}$ as $h(K)$, where $K$ is the positive number of opportunities. Since the choice of opportunity $m$ is rational (as opposed to purely probabilistic), $\varepsilon_{qjm}$ follows statistical distribution that generally differs from that of $\varepsilon_{qjk}$. It is convenient to focus on the limiting distribution of $h(K)$: $\lim_{K\to\infty} h(K)$. Under Gnedenko conditions (Kotz and Nadarajah 2000, p. 6), $h(K)$ converges in distribution to an extreme value distribution. For a normally distributed $\varepsilon_{qjk}$, the limiting distribution of $\varepsilon_{qj} \equiv \varepsilon_{qjm}$ is the type I extreme value distribution. The choice of opportunities does not affect other aspects of decision-making and thus can be conveniently omitted from further analysis. The important implication of the above analysis is that stochastic components $\varepsilon$ are independently identically distributed and drawn from the type I extreme value distribution with the joint probability density function

$$f(\varepsilon_{11}, \ldots, \varepsilon_{nm}) = \prod_{q=1}^{n} \prod_{j=1}^{m} \exp(-e^{-\varepsilon_{qj}})e^{-\varepsilon_{qj}}. \tag{3}$$

Using the type I extreme value distribution in (3) is sensible because it is the limiting distribution for many exponential family distributions such as normal, exponential, and logistic. For exotic underlying distributions such as Cauchy, log-normal, and uniform, the limiting distribution can be easily transformed to the type I extreme value distribution, so the case in (3) can be easily extended to a wide range of distributions.

The combination of the spatial random utility model (1), the decision rule (2), and the probability measure (3) comprises the basic spatial discrete choice model.

# 4 Likelihood-Based Estimation

## 4.1 Pseudo Maximum Likelihood Estimator

Identification of the spatial discrete choice model for a likelihood-based estimation varies to a large extent with the structure of observed utility $v_j(\beta)$ (for instance, multinomial logit vs conditional model). The common aspect of identification is, however, the conditions associated with spatial interactions. In this respect, it is important to mention three of them: (1) bounded spatial autoregressive coefficient $\rho$ so that the matrix $I - \rho W$ is a positive definite; and (2) limitations on topology of the non-zero elements of the spatial weights matrix (Smirnov and Anselin 2009) in order to satisfy asymptotic identification properties for spatial autoregressive models (Lee 2004); and (3) the absence of disconnected individuals. In the analysis below, the model is assumed to be identified and asymptotically identified.

The log-likelihood function for the discrete choice model (1) – (3) is

$$L(\theta; y) = \ln P(y \big| \beta, \rho), \tag{4}$$

where $P(y \big| \beta, \rho) = \text{Prob}(Y_{1k(1)} = 1, Y_{2k(2)} = 1, \dots, Y_{nk(n)} = 1 \big| \beta, \rho)$ is the joint probability for the discrete random variable $Y$ taken at $y$ and $k(q)$ is the alternative chosen by the agent $q$. Denote indicator function $I(a) : \mathbf{R} \to \{0, 1\}$ which takes value 1 if $a$ is positive and zero otherwise. The joint probability for (4) is given by

$$P(y \big| \beta, \rho) = \int \dots \int \Big( \prod_{q=1}^{n} \prod_{i=1}^{m} I(u_{qk(q)} \geq u_{qi}) \Big) f_\varepsilon(\varepsilon_{11}, \dots, \varepsilon_{nm}) d\varepsilon, \tag{5}$$

where $d\varepsilon = d\varepsilon_{11} \dots d\varepsilon_{1m} \dots d\varepsilon_{n1} \dots d\varepsilon_{nm}$. Using joint probability to obtain the likelihood function and setting up the likelihood maximization problem, one obtains the maximum likelihood estimator. However, the term $(I - \rho W)^{-1} \varepsilon$ is analytically intractable (Anselin 2002), which precludes the analytical formulation of the likelihood in (4). The conventional approach involving the calculation of $P(y|\beta,\rho)$ for any $(\beta,\rho) \in \Theta$ is an extremely challenging task because the integral in (5) cannot be factored into a product of easier-to-compute integrals of smaller dimensions, as typically is accomplished in models without spatial interactions between individuals (Train 2003; Sivakumar and Bhat 2007). The sheer dimension of the computational problem makes it impractical if not impossible to evaluate (4) using numerical integration.

To obtain the reduced form for the spatial random utility (1), one collects terms with $u$ on the left-hand side of the equation and premultiplies the result with the non-singular matrix $(I - \rho W)^{-1}$. The reduced form of the spatial random utility model (1) is

$$u_j = Zv_j(\beta) + Z\varepsilon_j, j \in \mathcal{M}, \tag{6}$$

where $Z = (I - \rho W)^{-1}$ is the spatial multiplier matrix. Matrix $Z$ is a non-singular positive definite matrix as implied from identification conditions. It is easy to show that it is non-negative, $Z_{ij} \geq 0$ for $\rho \geq 0$, and that its main diagonal exceeds identity matrix, $Z_{jj} \geq 1$ for $\rho \in \Theta_\rho$. The proof follows from the analysis of the convergence of the Taylor series

$$(I - \rho W)^{-1} = \lim_{n\to\infty} I + \rho W + \rho^2 W^2 + \cdots + \rho^n W^n \tag{7}$$

and non-negativity of the spatial weights matrix $W$. Another important property of the spatial multiplier matrix follows from the consideration that for an irreducible matrix $W$ and positive $\rho$, all elements of $Z$ are strictly positive because $W^n$ is a strictly positive matrix for some $n \leq N$.

Elements of matrix $Z$ indicate spatial multiplier effects; i.e., $Z_{qt} = \partial u_{qj}/\partial \varepsilon_{tj}$ – the full effect of random shock in the utility of individual $t$ on the random utility of individual $q$. As follows from (7), the immediate non-spatial effect of random shock $\varepsilon_{tj}$ equals $\varepsilon_{tj}$. The first-order spatial effect affects first-order neighbors of $t$ and equals $\rho w_{qt}$ for the individual $q$. The second-order spatial effect of $t$ on $q$ is the aggregate effect mediated by third parties, and equal to $\rho^2(W^2)_{qt}$, and so on. Every element of the spatial multiplier matrix $Z$ is the aggregate of the progressively discounted sequence of primary, secondary, tertiary, and so on effects that are mediated respectively by zero, one, two, etc., individuals. As series (7) indicates, for any positive $\rho$, $Z \geq I + \rho W$, hence, the full effect of spatial interactions that is given by the spatial multiplier matrix is larger than the immediate ($I$) and first-order ($\rho W$) spatial effects.

Denote $n \times n$ matrix $D$ that consists of diagonal elements of the matrix $Z$. The diagonal matrix $D$ indicates private effects of random shocks on the individual utilities. As follows from (7), these effects are the sum of direct non-spatial effects and aggregate spatial effects. Direct non-spatial effects are given by the identity matrix $I$ – the immediate effect of a local shock $\varepsilon_{qj}$ is the shock itself. Aggregate spatial effects are given by the matrix $Z - D$, which indicates the full spatial effect of a shock in the individual utility on the utilities of other individuals.

The conditional choice probability for the individual $q$ to select an alternative $j$ is

$$P_{qj} = P(y_{qj} = 1 \big| \{\varepsilon_{si}, s \in (\mathcal{N} \setminus q), i \in \mathcal{M}\}, \theta). \tag{8}$$

From (6), the random utility is

$$u_j = Zv_j(\beta) + Z\varepsilon_j = Zv_j(\beta) + (Z - D)\varepsilon_j + D\varepsilon_j. \tag{9}$$

Notice that the diagonal elements in the matrix $Z - D$ are zero, thus the vector of conditional choice probabilities is

$$P_j = Prob(\prod_{i \neq j} I(Zv_j(\beta) + (Z - D)\varepsilon_j \geq Zv_i(\beta) + (Z - D)\varepsilon_i), \tag{10}$$

which implies individual conditional probabilities

$$P_{qj} = \exp(g_{qj}/d_{qq}) \Big/ \sum_{i=1}^{m} \exp(g_{qi}/d_{qq}), \tag{11}$$

where

$$g_{qj} = \sum_{t=1}^{n} z_{qt} v_{tj}(\beta) + \sum_{t=1}^{n} (z_{qt} - d_{qt})\varepsilon_{tj}.$$

Notice that random components $\varepsilon_{tj}$ in $g_{qj}, j \in \mathcal{M}$ are independently identically distributed across alternatives for each individual, that is

$$E[\sum_{t=1}^{n} (z_{qt} - d_{qt})\varepsilon_{tj}] = c_q, \forall j \in \mathcal{M}, \tag{12}$$

and $Var[g_{qj} g_{qi}] = 0, i \neq j$. This indicates that the effect on the individual $q$'s utility conveyed by the spatial multiplier component $z_{qt}, t \neq q$ is not systematic and, hence, has no systematic effect on the conditional choice probability $P_{qj}$.

The major difficulty of the maximum likelihood estimation is the randomness of conditional probabilities $P_{qj}$ in (8). Suppose random shocks in utilities do occur according to the model, but individuals simplify their decision-making by focusing on spatial effects that systematically affect conditional choice probabilities and disregarding all other effects. It is easy to establish by examination of (11) and (12) that private shock $z_{qq} \equiv d_{qq}$ always affects conditional choice probability, whereas $z_{qs}, q \neq s$ has zero expected effect on individual random utility. The private effect of shock $\partial u_{qj}/\partial \varepsilon_{qj} = z_{qq}$ is always positive, identical across alternatives, $j \in \mathcal{M}$, but varies across individuals. It is important for any individual because $z_{qq}$ inadvertently affects conditional choice probabilities in a systematic way as further shown below in (13).

The homogeneity of spatial effects $g_{qj}$ on random utilities across alternatives suggests that spatial effects of shocks in random utilities of others are expected to have zero effect on conditional choice probabilities. For this reason, individuals might choose to disregard shocks that have no systematic effects on individual choices. Disregarding nuisance effects by no means negates the entire model of spatial interactions, because in non-interactive classical random utility models, random utility shocks are typically assumed to be synonymous with shocks in individual utility: $\partial u_{qj}/\partial \varepsilon_{qj} = 1$. The relevant repercussion of the spatial random utility model is that the presence of interactions alone introduces 'individuality' to the model, hence, individual differences in response to the shock. In addition, it is easy to show that for $\rho \in (0; 0.5)$, the diagonal element of the multiplier matrix is the largest element in the row for every individual regardless of the spatial arrangement. That is, the spatial effect from a shock has the strongest effect on the individual to which that shock is attributed. Hence, an individual is more concerned with the private effect of random shock than with the effects of shocks in other individuals' utilities. That said, the simplified closed form for the frequency estimator for conditional probabilities is

$$\hat{P}_{qj} = \exp(\sum_{t=1}^{n} z_{qt} v_{tj}(\beta)/d_{qq}) \Big/ \sum_{i=1}^{m} \exp(\sum_{t=1}^{n} z_{qt} v_{tj}(\beta)/d_{qq}). \tag{13}$$

Unlike $P_{qj}$, which is infeasible because $\varepsilon_{qj}$ are unobserved, the estimator $\hat{P}$ is feasible. Substituting $\hat{P}_{qj}$ for $P_{qj}$ in the log-likelihood function yields

$$\mathcal{L}(\beta, \rho | y) = \sum_{q=1}^{n} \sum_{j=1}^{m} y_{qj} \log \hat{P}_{qj}, \tag{14}$$

and the pseudo maximum likelihood estimator of $\beta$ and $\rho$ is the extremum estimator $\hat{\theta}_{PML} = (\hat{\beta}_{PML}, \hat{\rho}_{PML})$ maximizing the log-likelihood function (14). Since the relevant information about dependencies between probabilities is discarded, but no distorting alterations in the likelihood undertaken, the pseudo maximum likelihood estimator $\hat{\theta}_{PML}$ is consistent, but does not need to be asymptotically efficient.

In sum, the PML estimator introduced in this section is equivalent to the maximum likelihood estimator for the spatial discrete choice model with spatial random utility

$$\tilde{u}_j = Z(\rho) v_j(\beta) + D(\rho) \varepsilon_j, j \in \mathcal{M}. \tag{15}$$

The auxiliary model comprised of (15), (2) and (3) has the same observed deterministic components of individual random utilities as the original model (1) – (3). Private components of the unobserved spatial interdependencies ($D(\rho)\varepsilon_j$) in random utilities are also identical in both models. However, error terms in the auxiliary model are independent, which substantially simplifies its maximum likelihood estimation. Since some of the information about effects of individual interdependencies is not present in the auxiliary model, parameter estimates obtained from the PML do not need to be asymptotically efficient.

## 4.2 Computational Issues

The computation of the PML estimator is much easier than full maximum likelihood or use of importance sampling as suggested in (Murdoch, Sandler, Vijverberg 2003) which is applicable to fairly small datasets. In contrast, the PML is easier to implement and substantially lifts the limitation on the size of datasets and the number of alternatives. Computation of the log-likelihood function in the PML estimation is straightforward provided matrices $Z$ and $D$ are computed.

Since the spatial weights matrix $W$ is row-standardized from a symmetric matrix $W_0$, it is useful to keep the square roots of sums of the rows of matrix $W_0$ as a diagonal matrix $S$. Then, $W = S^{-2}W_0$, and $W_s = SWS^{-1}$ are similar, but $W_s$ is symmetric.

The computation of $(I - \rho W)^{-1}$ is simplified by using the identity $(I - \rho W)^{-1} = S^{-1}(I - \rho W_s)^{-1}S$. Taking into account that $W_s$ is sparse for large datasets, the inverse of $I - \rho W_s$ can be obtained by sparse Choleski decomposition $I - \rho W_s = LL'$, inverting the factor matrix $M = L^{-1}$, and computing the required matrix $(I - \rho W)^{-1} = M'M$. In this scenario, the elements of matrix $D$ are obtained as the by-product of computing $(I - \rho W)^{-1}$. Since the sparse matrix factorization routines are gaining popularity in commercial and open-source packages, this path seems to be the easiest for obtaining desired matrices. Some caution, however, is needed in order to address the issue of numerical accuracy. As Rue (2001) noted, reordering sparse matrices in order to preserve sparsity in factorization routines affects the numerical accuracy of the inverse. Consequently, additional conditioning is needed to improve the accuracy of the inverse and, hence, the estimates.

To remove ambiguity associated with the accuracy of matrix factorization routines, the conjugate gradient method was deployed in the implementation of the PML. The conjugate gradient method solves $(I - \rho W)z = v_j(\beta)$ for $z$ without inverting or factoring the matrix $I - \rho W$. It is a finite iterative method that requires little memory for storing intermediate results – since the matrix is not factored, there is no problem of storing dense matrices – and computations can be performed easily and quickly in a desktop or mobile computing environment even for large models. Elements of the diagonal matrix $D$ are computed using the sparse conjugate gradient method for effective computation of the vector $z$ for a sparse vector $v$ (Smirnov 2005) that benefits from the fact that only a small fraction of intermediate vectors are updated in each iteration. Both techniques have been tested for accuracy and used in lieu of matrix factorization in the Monte Carlo study described below.

# 5 Performance of the PML estimator: an illustration using simulated data

To illustrate finite sample properties of the PML method, consider the spatial discrete choice model with random utility

$$u_j = \rho W u_j + x_{tj}\beta_1 + x_j\beta_2 + \varepsilon_j, \tag{16}$$

where alternative $j$ indicates travel alternative, $x_{qj}$ indicates travel cost for the individual $t$ using alternative $j$, and $x_j$ is the attribute of alternative $j$, which is uniform for all individuals. A Monte Carlo study was conducted using simulated spatial weights matrices and spatial data (variables $x_{qj}, x_j$, and $y_{qj}$). Model parameters are estimated using the PML method introduced above, and thus obtained estimates are compared to the true parameter values.

In the Monte Carlo study, individuals are associated with unique locations. The neighborhood structure is given by the spatial weights matrix $W$ that is exogenous to the model. The spatial arrangement of individuals is assumed to follow the regular hexagonal pattern characteristic to the central place theory. Hexagons with common boundaries are assumed to be neighbors. This allows each location to have up to six neighbors. To randomize the simulated grid, a fixed fraction of neighborhood ties (about thirty percent of all pairs) are severed. However, it is assured that each individual on the boundary of the simulated grid has at least one neighbor, and each individual located in the interior of the grid has at least two neighbors. The resulting lattice resembles a lattice emerging from the hexagonal grid, yet random. Initially, the contiguity-based spatial weights matrix $W_0$ is symmetric and binary. It is further row-standardized, and in such form used in the computations. The size of dataset corresponds to the dimension of the spatial weights matrix in each simulation. The dimensions of the simulated datasets are reported along the results of simulations.

TABLE 1 ABOUT HERE

Parameters $\beta$ for the model (16) are set $\beta_1 = -4$ and $\beta_2 = 2$ in all simulations. The coefficient $\rho$ varies by trials as specified in tables below and takes values 0.05; 0.20 and 0.50. The number of alternatives is 8 in all trials. Spatial variables are generated as follows. Exogenous variables are independently drawn from uniform distributions. Variable $\varepsilon$ is drawn from the type I extreme value distribution. Random utility is computed by applying the conjugate gradient method (Smirnov 2005) to the reduced form random utility

$u_j = (I - \rho W)^{-1}(x_{tj}\beta_1 + x_j\beta_2 + \varepsilon_j)$ for each alternative $j$. Discrete choice variable $y$ is computed according to the decision rule, so that $y_{qi}$ is assigned value 1 if $u_{qi} \geq u_{qj}, j = 1, 2, \ldots, 8$ and zero otherwise. In the case of a tie, one of the alternatives would have been randomly selected, but ties did not happen. The PML uses exogenous variables $x_t$ and $x$, discrete choice variable $y$, and exogenous spatial weights matrix $W$.

As shown in Table 1, the smallest bias and the smallest MSE are reported for estimates of the spatial autoregressive coefficient $\rho$. The bias of the estimate of $\beta_2$ is approximately half the magnitude of that for the estimate of $\beta_1$. The MSE for estimates of the coefficient $\beta_2$ are approximately one-quarter of those for estimates of $\beta_1$. This result is expected since the true value of $\beta_1$ is twice of $\beta_2$. The fact that $x_t$ varies across individuals and alternatives whereas $x$ varies only across alternatives is irrelevant. However, the estimates of $\rho$ are fairly accurate, notwithstanding a relatively small size of the spatial dataset.

The general trend for estimates of $\beta$, but not $\rho$, is that the larger $\rho$, the larger biases and MSE on estimates of $\beta$. This characteristic of the method is also expected, since the loss of information from ignoring social effects in random components of individual utilities makes estimates less efficient, and that loss increases with $\rho$. Apparently, the loss of efficiency does not affect so significantly estimates of $\rho$.

TABLE 2 ABOUT HERE

Table 2 lists the results of the experiments with datasets of various sizes, so that one can verify the consistency of the method. Indeed, MSE of parameter estimates for larger datasets are generally lower, as expected. Still, bias and MSE for estimates of the spatial autoregressive coefficient $\rho$ are the smallest among the estimates. This suggests that the loss of efficiency of the estimates of $\rho$ tends to be negligible for simulated data.

# 6 Conclusion

Spatial dependence between individuals in this paper is understood as the spatial dependence between individual preferences. In all other respects decision-makers are independent and their choices have no social effects. This setting allows development of the spatial random utility model, which can be viewed as an extension of the random utility model to the case of spatially interdependent individuals. This extension is relevant for developing applied discrete choice models using spatial data. A consistent pseudo maximum likelihood estimator for the model is developed, and its properties are illustrated using simulated data.

The main advantages of the proposed pseudo maximum likelihood method are in its consistency, ease of computation, and analytical tractability of the concept of maximum likelihood. The PML estimator is equivalent to the maximum likelihood estimator for the auxiliary model, where individuals fully account for the spatial multiplier effect in the observed variables, but account only for private effects of spatial interdependencies in the unobserved shocks in random utilities. This understanding allows the interpretation of the likelihood-based statistical inference results along with understanding the limitations of the approach.

The computationally feasible implementation of the estimation method does not involve matrix factorization, thus allowing for the application of the estimator to large models. Finite sample and asymptotic properties of the PML are illustrated using Monte Carlo experiments. It is shown that the estimator is consistent for the type of spatial data generated. The sample-based bias and mean-squared error of the non-spatial coefficient estimates have approximately linear convergence, and the MSE, but not the bias of the spatial autoregressive coefficient, also converges linearly. The average bias of the spatial autoregressive coefficient is fairly small regardless of the size of the dataset.

# References


[1] Aguirregabiria, Victor; and Pedro Mira (2007). "Sequential estimation of dynamic discrete game," *Econometrica, 75(1)*, (Jan., 2007), p. 1–53.

[2] Anselin, Luc (2002). "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models," *Agricultural Economics, 27(3)*, (Nov., 2002), p. 247–267.

[3] Anselin, Luc (2006). "Spatial Econometrics," In *T.C. Mills and K. Patterson (Eds.), Palgrave Handbook of Econometrics: Volume 1, Econometric Theory.* Basingstoke, Palgrave Macmillan, pp. 901–69.

[4] Bekhor, Shlomo; and Joseph N. Prashker (2008). "GEV-based destination choice models that account for unobserved similarities among alternatives," *Transportation Research, Part B, (42)*, p. 243–262.

[5] Ben-Akiva, Moshe E.; and Lerman, Steven R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press Series in Transportation Studies, 9, MIT Press, Cambridge, MA.

[6] Besag, Julian (1974). "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society. Series B, 36(2)*, p. 192–236.

[7] Bhat, Chandra R.; and Jessica Guo (2004). "A mixed spatially correlated logit model: formulation and application to residential choice modeling," *Transportation Research, Part B, (38)*, p. 147–168.

[8] Brock, William A.; and Steven N. Durlauf (2001). "Discrete choice with social interactions," *Review of Economic Studies, 68(2)*, p. 236–260.

[9] Brock, William; and Steven N. Durlauf (2002). "A multinomial choice model with neighborhood effects," *American Economic Review, 92*, p. 298–303.

[10] Brock, William A.; and Steven N. Durlauf (2007). "Identification of binary choice models with social interactions," *Journal of Econometrics, 140*, p. 52–75.

[11] Case, Anne C. (1992). "Neighborhood Influence and Technological Change," *Regional Science and Urban Economics, 22(3)*, (Sept., 1992), p. 491–508.

[12] Chakravati, Dipankar; Atanu R. Sinha; and Jaewhan Kim (2005). "Choice Research: A Wealth of Perspectives," *Marketing Letters, 16(3:4)*, (Dec., 2005), p. 173–182.

[13] Conley, Timothy G.; and Giorgio Topa (2007). "Estimating Dynamic Local Interactions Models," *Journal of Econometrics, 140(1)*, (Sep., 2007), p. 76–96.

[14] Dormann, Carsten F. (2007). "Assessing the validity of autologistic regression," *Ecological Modeling, 207*, p. 234–242.

[15] Dugundji, Elenna R.; and László Gulyás (2008). "Sociodynamic Discrete Choice on Networks in Space: Impacts of Agent Heterogeneity on Emergent Outcomes," *Environment and Planning B: Planning and Design, 35(6)*, p. 1028–1054.

[16] Dugundji, Elenna R.; and Joan L. Walker (2005). "Discrete Choice with Social and Spatial Network Interdependencies," *Transportation Research Record, 1921*, p. 70–78.

[17] Fleming, Mark M. (2005). "Techniques for Estimating Spatially Dependent Discrete Choice Models," in Anselin, Luc E.; Florax, Raymond J. G.

M.; and Rey, Sergio J. (Eds.) *Advances in Spatial Econometrics*, Springer-Verlag, Berlin, p. 145–168.

[18] Graham, Bryan S. (2008). "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica, 76(3)*, (May, 2008), p. 643–660.

[19] Iglesias, Emma M.; and Garry D. A. Phillips (2008). "Asymptotic Bias of GMM and GEL under Possible Nonstationary Spatial Dependence," *Economics Letters, 99*, p. 393–397.

[20] Klier, Thomas; and Daniel P. McMillen (2008). "Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples," *Journal of Business and Economic Statistics, 26(4)*, (Oct., 2008), p. 460–471.

[21] Koppelman, Frank S.; and Vaneet Sethi (2005). "Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior," *Transportation Research, Part B (39)*, p. 825–853.

[22] Kotz, Samuel, and Saralees Nadarajah (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London, UK.

[23] Lee, Lung-Fei (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica, 72(6)*, p. 1899–1925.

[24] Louviere, Jordan; Kenneth Train; Moshe Ben-Akiva; Chandra Bhat; David Brownstone; Trudy Ann Cameron; Richard T. Carson; J.R. Deshzo; Denzil Fiebig; William Green; David Hensher; and Donald Waldman (2005). "Recent progress on endogeneity in choice modeling", *Marketing Letters, 16(3:4)*, (Dec., 2005), p. 255–265.

[25] Louviere, Jordan J. (2006). "What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments," *Environmental and Resource Economics, 34(1)*, (May, 2006), p. 173–188.

[26] Manski, Charles (2004). "Measuring Expectations," *Econometrica, 72(5)*, (Sept., 2004), p. 1329–1376.

[27] McFadden, Daniel (1978). "Modeling the choice of residential location," in: *Karlqvist, A. et al. (Eds.), Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, p. 75–96.

[28] McMillen, Daniel P. (1992). "Probit with Spatial Autocorrelation," *Journal of Regional Science, 32(3)*, p. 335–348.

[29] Mohammadian, Abolfazl (Kouros); Murtaza Haider; and Pavlos S. Kanaroglou (2005). "Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models," *84-th Annual Transportation Research Board Meeting,* January 2005, Washington, D.C.

[30] Murdoch, James C.; Todd Sandler; and Wim P. M. Vijverberg (2003). "The participation decision versus the level of participation in an environmental treaty: a spatial probit analysis," *Journal of Public Economics, (87)*, p. 337–362.

[31] Páez, Antonio; and Darren M. Scott (2007). "Social Influence on Travel Behavior: A Simulation Example of the Decision to Telecommute," *Environment and Planning A, 39(3)*, p. 647–665.

[32] Páez, Antonio; Darren M. Scott; and Erik Volz. (2008). "A Discrete-choice Approach to Modeling Social Influence on Individual Decision Making," *Environment and Planning B: Planning and Design, 35(6)*, p. 1055–1069.

[33] Pinkse, Joris; and Margaret E. Slade (1998). "Contrasting in Space: An Application of Spatial Statistics to Discrete-Choice Models," *Journal of Econometrics, 85(1)*, p. 125–154.

[34] Pinkse, Joris; Margaret Slade; and Lihong Shen (2006). "Dynamic Spatial Discrete Choice Using One-step GMM: An Application to Mine Operating Decisions," *Spatial Economic Analysis, 1(1)*, (June 2006), p. 53–99.

[35] Rue, Håvard (2001). "Fast Sampling of Gaussian Markov Random Fields," *Journal of the Royal Statistical Society, Series B, 63*, p. 325–338.

[36] Sivakumar, Aruna; and Chandra R. Bhat (2007). "A Comprehensive, unified, framework for analyzing spatial location choice," In: *Proceedings of the 86th Annual Meeting of the Transportation Research Board,* Washington, DC.

[37] Smirnov, Oleg (2005). "Computation of the Information Matrix for the Models of Spatial Interaction on a Lattice," *Journal of Computational and Graphical Statistics, 14(4)*, p. 910–27.

[38] Smirnov, Oleg A., and Luc E. Anselin (2009). "An $O(N)$ Parallel Method of Computing the Log-Jacobian of the Variable Transformation for Models with Spatial Interaction on a Lattice," *Computational Statistics and Data Analysis, 53(8)*, (Jun., 2009), p. 2980–2988.

[39] Smith, Tony E.; and James P. LeSage (2004). "A Bayesian probit model with spatial dependencies," in James P. Lesage and R. Kelley Pace (Eds.) *Spatial and Spatiotemporal Econometrics, Volume 18*, Elsevier, Amsterdam, The Nethelands, p. 127–160.

[40] Train, Kenneth E. (2003). *Discrete Choice Methods with Simulations*, Cambridge University Press, New York, NY.

[41] Walker, Joan L.; and Jieping Li (2007). "Latent lifestyle preferences and household location decisions," *Journal of Geographical Systems, 9(1)*, (Apr., 2007), p. 77–101.

[42] Ward, Michael D.; and Kristan Skrede Gleditsch (2002). "Location, location, location: An MCMC approach to modeling the spatial context of war and peace," *Political Analysis, 10(3)*, p. 244–260.

Table 1: Accuracy of the parameter estimates for various values of the spatial autoregressive coefficient

| Number | $\rho$ | | $\beta_1$ | | $\beta_2$ | |
|---|---|---|---|---|---|---|
| of trials | bias | MSE | bias | MSE | bias | MSE |
| | | | $\rho = 0.05$ | | | |
| 5 | -0.00742 | 0.00037 | 0.17337 | 0.28320 | 0.08812 | 0.07891 |
| 20 | 0.00073 | 0.00012 | 0.10300 | 0.13047 | 0.05201 | 0.03190 |
| 60 | 0.00081 | 0.00020 | 0.07075 | 0.15393 | 0.02108 | 0.03996 |
| 100 | -0.00187 | 0.00014 | 0.07116 | 0.13738 | 0.02108 | 0.03796 |
| | | | $\rho = 0.20$ | | | |
| 5 | -0.00217 | 0.00006 | -0.01056 | 0.10124 | 0.00400 | 0.03147 |
| 20 | 0.00204 | 0.00163 | 0.03366 | 0.19030 | 0.02652 | 0.05483 |
| 60 | 0.00094 | 0.00022 | 0.11515 | 0.18694 | 0.04805 | 0.04685 |
| 100 | -0.00229 | 0.00021 | 0.05747 | 0.18281 | 0.03739 | 0.05211 |
| | | | $\rho = 0.50$ | | | |
| 5 | -0.00132 | 0.00010 | -0.3751 | 0.22809 | -0.13902 | 0.04411 |
| 20 | -0.00047 | 0.00010 | 0.03513 | 0.12612 | 0.02248 | 0.04283 |
| 60 | -0.00032 | 0.00007 | -0.10852 | 0.16989 | -0.04745 | 0.04234 |
| 100 | 0.00212 | 0.00009 | -0.11730 | 0.12115 | -0.06224 | 0.03714 |

*Note: Spatial dataset contains 400 locations. Bias is the average bias of the estimate over a given number of trials. MSE is the mean squared error of the estimator.*

Table 2: Accuracy of the method depending on the size of the spatial dataset

| Number of trials | $\rho$ bias | $\rho$ MSE | $\beta_1$ bias | $\beta_1$ MSE | $\beta_2$ bias | $\beta_2$ MSE |
|---|---|---|---|---|---|---|
| | | | size of the dataset is $N = 400$ | | | |
| 5 | -0.00217 | 0.00006 | -0.01056 | 0.10124 | 0.00400 | 0.03147 |
| 20 | 0.00204 | 0.00163 | 0.03366 | 0.19030 | 0.02652 | 0.05483 |
| 60 | 0.00094 | 0.00022 | 0.11515 | 0.18694 | 0.04805 | 0.04685 |
| 100 | -0.00229 | 0.00021 | 0.05747 | 0.18281 | 0.03739 | 0.05211 |
| | | | size of the dataset is $N = 5,000$ | | | |
| 5 | -0.00130 | $4 \cdot 10^{-6}$ | 0.00710 | 0.02567 | -0.00616 | 0.00841 |
| 20 | -0.00165 | 0.00002 | -0.02555 | 0.01103 | -0.00489 | 0.00330 |
| 60 | 0.00029 | 0.00003 | -0.02233 | 0.01702 | -0.01164 | 0.00549 |
| 100 | 0.00006 | 0.00003 | -0.01890 | 0.02007 | -0.01098 | 0.00605 |
| | | | the size of the dataset is $N = 20,000$ | | | |
| 5 | 0.00190 | 0.00001 | 0.03612 | 0.00896 | 0.00669 | 0.00160 |
| 20 | -0.00027 | $7 \cdot 10^{-6}$ | 0.00662 | 0.00625 | 0.00386 | 0.00174 |
| 60 | -0.00070 | $8 \cdot 10^{-6}$ | -0.01472 | 0.00654 | -0.00566 | 0.00165 |
| 100 | -0.00041 | $7 \cdot 10^{-6}$ | -0.00499 | 0.00494 | -0.00005 | 0.00139 |

*Note: In all trials* $\rho = 0.20$.