A GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL DISCRETE-CHOICE MODELS INVOLVING MICRO-LEVEL DATA

By

LEDIA GUCI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To my family and the memory of my uncle, Terry A. Terezi

ACKNOWLEDGMENTS

The completion of this dissertation was made possible with valuable contribution by many individuals so I would like to take this opportunity to thank them all. First and foremost, I wish to extend my deepest gratitude to my dissertation committee chair, Dr. Alfonso Flores-Lagunes; his guidance and encouragement went well beyond the mere supervision of a research project. His high standards made me learn to appreciate the rigor of research and find the motivation needed to accomplish this task. I am especially indebted to Dr. Flores-Lagunes for assisting in shaping my career path by providing me with outstanding mentorship and numerous professional opportunities.

I am also utterly grateful to my other supervisory committee members, Dr. Charles Moss, Dr. Carmen Carrion-Flores, and Dr. Dawn Jourdan for their constructive criticism regarding various aspects of the project which encouraged me to broaden my research perspective. In addition, I am very thankful to Dr. Ray Huffaker for his involvement and recommendations on earlier stages of this work, Dr. David Brasington of the University of Cincinnati for compiling and supplying the data set used in one of the empirical studies, and James Colee, a consultant with the IFAS Statistics Department, for his suggestions regarding the generation of the data for some of the Monte Carlo experiments. I feel fortunate to have collaborated with and learned from such dedicated scholars. However, it goes without saying that I am solely responsible for any errors found in this manuscript.

I would also like to express my appreciation to the Food and Resource Economics Department for the financial support that made my graduate studies possible. Special thanks go to Mrs. Jessica Herman for her help with the graduate program logistics and

my professors and fellow graduate students for making this pursuit of knowledge a memorable experience.

Finally, and most importantly, I wish to extend my heartfelt thanks to my family and friends for providing the much needed moral support as they patiently waited for the completion of this work. Special appreciation is offered to my grandmother, my parents, my sisters, and my fiancé for sharing the burden as well as the excitement of this experience. I can only hope that I have made you proud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDF	Cumulative distribution function (of a random variable)
CL	Conditional logit (model)
CLR	Classical linear regression (model)
FGLS	Feasible generalized least squares (estimation)
GLS	Generalized least squares (estimation)
GMM	Generalized method of moments (estimation)
IIA	Independence of irrelevant alternatives (property)
IID	Independent and identically distributed (random variable)
IV	Instrumental variable (estimation)
LM	Lagrange multiplier (test)
LR	Likelihood ratio (test)
ML	Maximum likelihood (estimation)
MNL	Multinomial logit (model)
MXL	Mixed (multinomial) logit (model)
OL	Ordered logit (model)
OLS	Ordinary least squares (estimation)
PDF	Probability density function (of a random variable)
RMSE	Root mean-squared error (statistic)
RUM	Random utility maximization (framework)
SAE	Spatial autoregressive error (model)
SAL	Spatial autoregressive lag (model)
SAL-SAE	Spatial autoregressive model with spatial autoregressive errors
SCL	Spatial conditional logit (model)
SMNL	Spatial multinomial logit (model)

SMXL	Spatial mixed logit (model)
SOL	Spatial ordered logit (model)
TSLS	Two-stage least squares (estimation)

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL DISCRETE-CHOICE MODELS INVOLVING MICRO-LEVEL DATA

By

Ledia Guci

August 2011

Chair: Alfonso Flores-Lagunes
Major: Food and Resource Economics

Many economic problems that require micro-level analysis within a discrete-choice framework are fundamentally spatial processes. Under these circumstances, the estimation of standard discrete-choice models results in inconsistent parameter estimates. While several estimation methodologies have been developed for spatial discrete-choice models, their implementation becomes infeasible in empirical applications involving micro-level data.

This dissertation extends an estimation methodology for discrete-choice models with spatial lag dependence that is computationally feasible with large samples. The methodology consists of linearizing the original spatial model around a convenient point of parameter values and estimating the linearized model using a generalized method of moments (GMM) approach. The model linearization avoids the repeated matrix inversion involved in GMM estimation – which hinders its use with large samples – and breaks up the estimation procedure into two steps: a standard discrete-choice model followed by two-stage least squares estimation. Spatial estimators based on this methodology are derived for various polychotomous logit models. The performance of the proposed methodology in finite samples is assessed using Monte Carlo methods.

Simulation results indicate that the linearized model provides a good approximation to the original spatial model for a reasonable range of induced spatial dependence in the simulated data.

The proposed spatial estimation approach is employed in two empirical studies. The first study uses parcel-level data to estimate a spatially-explicit model of land-use conversion occurring at the rural-urban fringe. The results from this study corroborate previous research findings regarding the tendencies of new urban development. Moreover, the results provide new evidence of significant spatial dependence in land-use conversion decisions. The presence of spatial spillover effects suggests that policies designed at a small scale could lead to sub-optimal land-use patterns. The second study examines school districts' adoption decisions of interdistrict open enrollment policies using a spatially-explicit model of policy adoption to accommodate for potential influence from neighboring districts. The empirical results substantiate extant descriptive evidence regarding the determinants of adoption of open enrollment policies. Importantly, the results show strong neighborhood influence in policy adoption decisions calling attention to the critical role that reference groups play in shaping decision makers' behavior toward the adoption of new policies.

CHAPTER 1
INTRODUCTION

The economic behavior of decision makers faced with discrete-choice decisions that involve various alternatives is often influenced by the behavior of decision makers in reference groups – neighbors, peers, or other individuals that face similar decisions – because it is costly to evaluate all alternatives. The economics literature has documented numerous settings that have prompted such interdependence in decision making. Examples include technology adoption decisions (e.g. Katz and Shapiro, 1986; Case, 1992), financial investment decisions (e.g. Scharfstein and Stein, 1990), firms' choice of sales contracts (Pinkse and Slade, 1998), landowners' land-use conversion decisions (e.g. Irwin and Bockstael, 2004; Zhou and Kockelman, 2008), firms' location decisions (Klier and McMillen, 2008), etc. Thus, modeling individual discrete-choice decisions often requires accommodating for interdependence in decision making, which involves specifying and estimating spatial discrete-choice models.

A spatial discrete-choice model formalizes the spatial relationships that exist between the decision makers by defining a spatial structure in the form of a spatial weights matrix that explicitly links each of the observations in the sample and assigns spatial weights based on the strength of the potential interaction between observations at different locations. In the presence of such spatial interactions, the estimation of standard discrete-choice models is problematic due to a number of econometric challenges such as spatial heteroskedasticity and autocorrelation, spatial heterogeneity, and selection bias. Thus, traditional modeling methods generally lead to inconsistent estimates and are inappropriate for hypothesis testing and prediction (McMillen, 1992).

The estimation of spatial discrete-choice models is challenging because of nonlinear optimization procedures involved in the estimation of discrete-choice models which are further complicated by spatially dependent observations. In addition, the sample size plays a critical role. The maximum likelihood procedure, typically employed in discrete-choice model estimation, becomes increasingly cumbersome with larger samples because the likelihood function involves numerous integrals under non-spherical disturbances. Generalized method of moments estimation becomes infeasible as well in large data sets because it involves iterative numerical optimization processes which require repeated inversion of large matrices.

Several spatial discrete-choice estimators in the literature provide consistent estimates under spatially dependent data. Their application, however, has been limited to the binary choice case and small sample sizes. These estimators attempt to preserve the estimation structure implied by maximum likelihood by either making simplifying assumptions about the spatial weights matrix or directly simulating the choice probabilities under a set of assumptions for the error terms (e.g. Case, 1992; LeSage, 2000; McMillen, 1992, Beron and Vijverberg, 2004). More recent attempts to incorporate spatial dependence in a multinomial setting follow a Bayesian framework (e.g. Kakamu and Wago, 2005; Wang and Kockelman, 2009; Chakir and Parent, 2009). These approaches are computationally intensive as well in moderate to large sample sizes.

This dissertation develops an estimation methodology for cross-sectional polychotomous choice models with spatial lag dependence that is computationally feasible in large samples. The methodology – an extension of the methodology developed by Klier and McMillen (2008) for the estimation of a spatial binary logit model

– consists of linearizing the original spatial model around a convenient point of initial parameter values and estimating the linearized version of the spatial model. The extension of Klier and McMillen's (2008) methodology is possible because the key result that enables the linear approximation of the spatial model in a binary choice case holds in a multinomial setting. The model linearization avoids the repeated matrix inversion involved in GMM estimation and breaks up the estimation procedure into two simple sequential steps: a standard polychotomous choice model with no spatial dependence followed by a two-stage least squares estimation of the linearized spatial model which accounts for the spatial dependence.

Spatial estimators based on this linearization methodology are derived for various polychotomous models, all of which have a wide applicability in analyzing economic decisions. Because these spatial estimators are based on linearized versions of the original spatial models, how well they perform depends on how well the former models approximate the latter. Thus, the performance of the spatial estimators in finite samples is assessed in a controlled setting using Monte Carlo methods. Simulation results indicate that the linearized spatial models provide a good approximation to the original spatial models and produce fairly accurate estimates for a reasonable range of induced spatial dependence in the simulated data. Judging from these results, the linearization approach appears successful. However, the ultimate assessment of the use and performance of the proposed estimation method rests in the empirical applications.

The proposed spatial estimation approach is used in two empirical studies. The first study involves a spatial analysis of landowners' land-use conversion decisions. The main objective of this study is to understand what drives land-use conversion and

identify those factors that play a significant role in the conversion of land to urban use. It uses parcel-level data from Medina County, Ohio, to estimate a spatially-explicit model of land-use conversion occurring at the rural-urban fringe. The focus of the second study is neighborhood influence in the adoption of public policies. For this study school district data are used to examine the neighborhood influence in the adoption of interdistrict open enrollment policies by school districts in Ohio. Both studies find significant spatial dependence in the discrete-choice decisions under consideration. Knowledge of these spatial effects is of value to inform policy.

The outline of this dissertation is as follows. Chapter 2 provides an overview of the literature on spatial linear regression models and discrete-choice models to lay the foundation necessary for the development of the estimation methodology proposed in the next chapter. Chapter 3 reviews the literature on spatial nonlinear models and discusses the challenges that arise when estimating discrete-choice models in the presence of spatial effects. It then proposes an estimation approach for a class of spatial multinomial logit models that is computationally simple. In Chapter 4 the proposed methodology is first assessed using Monte Carlo experiments and then used to empirically estimate a spatially-explicit model of land-use conversion. Chapter 5 examines the neighborhood influence on school districts' adoption decisions of interdistrict open enrollment policies using a spatially-explicit model of policy adoption. The last chapter provides final remarks and outlines venues of future work.

CHAPTER 2
SPATIAL LINEAR REGRESSION MODELS AND DISCRETE-CHOICE MODELS: A LITERATURE REVIEW

Spatial Linear Regression Models

Spatial Dependence and Spatial Heterogeneity

Spatial effects are present in various empirical applications in which observations in the data sample are arranged in space[1]. Prominent examples include empirical applications that utilize data on population, employment, or other economic activities recorded over administrative units such as states, counties, districts, etc. Based on spatial features of the data, the spatial econometric literature distinguishes between two types of spatial effects: spatial dependence and spatial heterogeneity[2].

Spatial dependence is defined as "the existence of a functional relationship between what happens at one point in space and what happens elsewhere (Anselin, 1988a, page 11)". Anselin and Bera (1998) define spatial dependence as the "coincidence of value similarity with location similarity". The values of a random variable exhibit a pattern over space as values associated with two locations that are closer tend to be more correlated – either positively or negatively – than the ones associated with locations that are further apart. Similar values of a random variable that tend to cluster in space are an indication of positive spatial dependence. In contrast, negative spatial dependence is present when a random variable takes on dissimilar values for

[1] The concept of space is not confined to geographic space. It can be generalized to include policy space, economic space, technological space, social networks, etc. Refer to Isard (1969) for further discussion.

[2] In the spatial econometrics literature, the terms 'spatial dependence' and 'spatial autocorrelation' are often used synonymously. Strictly speaking, spatial dependence implies a multidimensional relationship of the observations in the sample which is expressed by a joint density function, whereas spatial autocorrelation is a process characterized by moments of this joint distribution. The latter, a weaker expression of the former, is the focus of most applications. Spatial dependence and spatial autocorrelation are used interchangeably throughout this manuscript.

observations in neighboring locations, resulting in dispersed patterns. Positive spatial dependence has a more intuitive interpretation and it is the kind of spatial dependence most commonly encountered in empirical applications. Other formal definitions of spatial dependence are also available in the literature (e.g. Griffith, 1992; Hepple, 2000).

Consider the classical linear regression (CLR) model:

$$Y_i = X_i\beta + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \tag{2.1}$$

where Y is the dependent variable; X denotes the explanatory variables; ε is a vector of stochastic errors; and β represents the model parameters. In this model, spatial dependence between two observations corresponding to locations i and j can be formally expressed as[3]:

$$Y_i = f(Y_j), \qquad \forall\, i \neq j, i = 1, \dots, N \tag{2.2}$$

Note that the observations Y_i and Y_j in (2.2) do not obey the structure of the data generating process of a cross-sectional data sample of N independent observations implied by the linear regression model in (2.1). Thus, the estimation of the model by means of ordinary least squares (OLS) generally leads to biased and inconsistent estimates and invalid statistical inference. These undesirable properties of the OLS estimator have led to the development of specialized estimation techniques that explicitly account for the structure of spatial dependence. A discussion of these estimation methods follows.

Spatial heterogeneity, on the other hand, is defined as the lack of structural stability (or uniformity) of spatial relationships over space (Anselin, 1988a). In this case, the spatial process differs across spatial units resulting in spatial effects that vary over

[3] In terms of the moment condition, spatial autocorrelation is formally defined as: $Cov(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j) \neq 0, \forall\, i \neq j$.

space. As a result, the estimated spatial effects over all locations may not adequately describe the spatial process at a particular location. A general case of spatial heterogeneity in a CLR model can be expressed as:

$$Y_i = X_i\beta_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \tag{2.3}$$

where the model parameters β_i imply the existence of a different linear relationship between the dependent variable and the explanatory variables for every observation in the sample[4]. In general, problems caused by spatial heterogeneity (e.g. spatial heteroskedasticity) can be addressed with standard econometric techniques (Anselin, 1988a). Thus, this review on spatial econometric models as well as the research presented in this dissertation focus exclusively on issues of spatial dependence.

In cross-sectional data, both types of spatial effects may appear alike. For instance, an observed spatial cluster of outliers in OLS residuals could be a case of spatial heteroskedasticity or it could be due to a spatial process that generated the residual cluster, thus a case of spatial autocorrelation. The problem of discerning between these effects is known in the literature as the problem of “true” contagion versus “apparent” contagion (Anselin and Bera, 1998). Moreover, spatial dependence and spatial heterogeneity may coexist. In this case distinguishing between the two types of spatial effects is especially complex. The spatial econometrics literature deals with these spatial effects by incorporating them explicitly in the model and testing for the specification of the spatial model[5].

[4] In this case, spatial heterogeneity violates the CLR model’s assumption of a single linear relationship between the dependent variable and the explanatory variables for all the observations in the data sample. It may also involve a different functional form of this relationship for different observations.

[5] This model-driven approach is what distinguishes the two closely related fields of spatial econometrics and spatial statistics. The latter adopts a data-driven approach in which randomness of spatial effects is assumed and the spatial patterns and interactions are derived from the data. In contrast, the model-driven

Spatial Weights

Analysis of spatial dependence requires an explicit expression of the spatial pattern which quantifies the extent and the strength of spatial interactions between spatial units (observations). A cross-sectional data sample with N observations contains insufficient information for the estimation of the parameters corresponding to all spatial interactions[6]. Thus, it is necessary to impose a structure on these interactions and express them as a function of a smaller number of parameters that can be estimated[7] (Anselin, 1988a). The structure of spatial dependence is formally expressed by defining a NxN spatial weights matrix W that links all the observations in the sample. This is done by assuming that spatial dependence arises within a "neighborhood" of data points surrounding the spatial unit of interest. Thus, the spatial weights matrix contains non-negative constant elements (w_{ij}) for those observations that are defined as neighbors based on some criterion that properly describes the nature of the interaction process. In addition, the spatial weights matrix W is symmetric and, by convention, the diagonal elements are set to zero $(w_{ii} = 0)$ indicating that an observation is not a neighbor to itself. Typically, the spatial weights are assumed to be exogenous and determined a priori.

---

approach starts with a theory and is eventually confronted with the data, but focuses on model estimation, specification testing, and statistical inference in the presence of spatial effects (Anselin, 1988a).

[6] Spatial dependence implies the existence of a NxN covariance (correlations) matrix; hence, it requires estimation of up to $N^2 - N$ parameters with N observations.

[7] The structure on the spatial interactions can be imposed based on two approaches. The geostatistical or "direct" approach models directly the correlations between the pairs of observations as a continuous function of the distance between their locations. The lattice or "spatial weights matrix" approach models the correlations indirectly from a particular spatial stochastic process specified to describe the interaction process. The lattice approach is more appropriate for economic analyses where economic agents are associated with discrete locations in space (Anselin, 2002). For this reason, the lattice approach is reviewed in this section and adopted in the empirical studies presented in the subsequent chapters.

The original form of the spatial weights matrix (by Moran, 1948; Geary, 1954) is based on the concept of binary contiguity between two spatial units; w_{ij} is set equal to one if two spatial units share a common border (i.e. are contiguous) and zero otherwise. This definition of contiguity assumes the existence of a map-like layout with distinguishable boundaries and works well for areal spatial units with irregular boundaries. When spatial units are organized in a regular lattice, a common border may be defined as a common edge, a common vertex, or a combination of both, resulting in different definitions of contiguity. These definitions of contiguity are known as the rook, the bishop, and the queen, respectively. Illustrations of the various definitions of contiguity are provided in Anselin (1988a). Spatial units may also consist of points (rather than areas) that are arranged in space. In this case, areal spatial units are generated by using various tessellations which partition the space into polygons that correspond to the location of the points[8]. In addition to simple contiguity, neighbors can be defined based on higher orders of contiguity; kth order contiguity results from spatial units that are first order contiguous to a kth-1 order contiguous spatial unit, but not contiguous to spatial units of lower order contiguity than $k-1$. In a regular lattice, neighbors based on higher orders of contiguity can be visualized as a series of concentric bands around the spatial unit of interest (Anselin, 1988a).

Another criterion commonly used to specify the spatial weights is based on the distance between two spatial units; w_{ij} is set equal to one if location i and j are contained within a specified distance threshold and zero otherwise. To measure the

[8] Thiessen polygons (also known as Dirichlet polygons or Voronoi polygons) are the most common types of spatial tessellations. See Ripley (1981), Amrhein et al. (1983), and Upton and Fingleton (1985) for an overview.

strength of the potential interaction between spatial units, spatial dependence is assumed to decrease with distance and the spatial weights are assigned according to an inverse function of distance; $w_{ij} = f(1/d_{ij})$, where d_{ij} is the distance between observations in locations i and j. Measures of distance can be defined based on geographical information (e.g. Euclidean distance, travel-time distance) as well as based on socio-economic indices. For instance, Case et al. (1993) defined "distances" between counties based on differences in per capita income and proportion of black population. Doreian (1980) specified spatial weights based on individuals' social networks.

More general spatial weights matrices combine distance and contiguity relationships. For instance, Cliff and Ord (1973, 1981) combined a distance measure with the length of the border shared between spatial units to capture the physical characteristics of the spatial unit. Their criterion results in an asymmetric spatial weights matrix with elements $w_{ij} = (d_{ij})^{-a}(\beta_{ij})^{b}$, where d_{ij} is the distance between location i and j, β_{ij} is the share of the boundary of location i contiguous to j, and a and b are parameters (set a priori). Similar weighing schemes can be found in Dacey (1968), Bodson and Peeters (1975), etc. Despite how the spatial weights are specified, the resulting spatial process must satisfy some regularity conditions in order to obtain desirable properties of estimators and test statistics [9](Anselin and Bera, 1998).

The spatial weights matrix is often standardized such that the sum of row elements equals one. Row-standardization is mainly done for statistical reasons so as to constrain the spatial autoregressive parameters to ensure that the likelihood function of

[9] These are known as the mixing conditions; the spatial weights must be nonnegative, finite, and generated by a proper metric. In addition, they have to be exogenous (Anselin, 1988a).

the spatial model is well-behaved over a non-continuous parameter space[10]. In addition, it ensures that all the spatial weights are between zero and one, which facilitates the interpretation of the spatial impacts as an "average" of neighboring values and makes the parameters of different spatial stochastic processes comparable between models (Anselin and Bera, 1998). Row-standardizing, however, may change the structure of the assumed interdependence because it generally rescales each row of the W matrix differently. As a result, the relative dependence among neighbors does not change, but the total impact of neighbors across observations changes (Bell and Bockstael, 2000). In this case, the economic interpretation of the total spatial impacts should be made with caution, especially for distance-decay type matrices, because the total impact of neighbors varies over observations depending on the density and the spatial pattern that surrounds the spatial unit (observation) of interest[11]. To avoid the interpretation problems from row-standardization of distance-decay spatial matrices, a series of higher order contiguity matrices of a more flexible form have been developed (Plantiga and Irwin, 2006). Row-standardization may also complicate the model estimation since the spatial weights matrix may become asymmetric after standardization.

In practice, a number of issues may arise when generating spatial weights based on any of the criteria described above. The contiguity criterion may result in different quantitative representations of the spatial relationships for the same spatial pattern. The distance criterion based on geographic information creates problems when there is a

[10] The likelihood function is well-defined over the parameter space when the spatial autoregressive parameters are constrained within the interval $1/\omega_{min}$ and $1/\omega_{max}$, where ω_{min} and ω_{max} are the smallest and the largest eigenvalues of W, respectively (Anselin and Bera, 1998).

[11] A discussion on the sensitivity of model results to both specification and row-standardization of W is provided in Plantiga and Irwin (2006) and Bell and Bockstael (2000).

high degree of heterogeneity in the distribution of spatial units or the size of spatial regions, because it results in a highly disproportionate number of neighbors for smaller spatial units compared to larger spatial units. Alternative measures of distance such as “economic” distance may produce weights that are not meaningful. For instance, a metric based on inverse distance produces weights that are undefined when two individuals have identical socio-economic indicators. In addition, the spatial weights are unlikely to be exogenous when the variables used to generate the weights are included in the model. Anselin (2002) provides a thorough discussion and provides examples for each of these issues.

It becomes evident from this discussion that the specification of the structure of the spatial dependence is critical to the model estimation. First, the specification of the spatial weights matrix is made a priori and it is not a testable. Once specified, the model estimation is carried out under the assumption that the spatial weights matrix represents the true spatial structure of the neighborhood. Thus, the estimates are valid as long as the spatial structure is correctly reflected in the spatial weights. Second, different specifications of the spatial dependence yield different covariance structures (Anselin, 2002). As a result, model estimates may be sensitive to the specification of the spatial weights matrix. For this reason, it has become customary for researchers to conduct sensitivity analyses by specifying several spatial weights matrices based on different metrics and neighborhood sizes.

For any particular empirical application, there is little formal guidance for the proper choice of the spatial weights matrix. This choice depends on the nature of the problem being modeled and any information on the problem that might be available.

Anselin (1988a) argues that the choice of spatial weights should have a direct link with the theoretical conceptualization of the structure of dependence, rather than describe the spatial pattern ad hoc. Proper specification of the spatial weights matrix, however, is complex and remains one of the most controversial methodological issues in the analysis of spatial data.

Spatial Lag Dependence and Spatial Error Dependence

In applied work, spatial dependence is generally an outcome of two conditions: measurement error or the existence of socio-economic phenomena that involve spatial interaction and result in spatial externalities and spatial spillover effects (Anselin, 1988a). Spatial dependence as a byproduct of measurement error arises when spatial data boundaries do not coincide with the spatial scale of the phenomenon under study, when aggregating spatial data, or when combining data samples defined at different spatial scales. As a result, errors from neighboring locations are correlated resulting in spatial autocorrelation. In this case, spatial dependence is considered a nuisance and the main objective of the spatial analysis is to obtain proper statistical inference. When spatial interaction characterizes the socio-economic phenomenon under study, the purpose of modeling spatial dependence is to understand the nature of the spatial interaction and the underlying economic and social processes that generated the interaction. In this case, spatial dependence is in itself of interest to explain economic behavior, hence considered a substantive. Spatial dependence is typically modeled by specifying a relationship between the dependent variable (Y) or the error term(u) and the corresponding spatial lags, $W_Y Y$ and $W_u u$. The resulting model specifications are known as the spatial autoregressive lag (SAL) and spatial autoregressive error (SAE) models, respectively.

Spatial dependence in the dependent variable, known as spatial lag dependence, is analogous to the inclusion of a serially autoregressive term of the dependent variable in a time-series model (Anselin and Bera, 1998). In this model, the value of the dependent variable corresponding to each cross-sectional unit is assumed to depend, in part, on the weighted average of the values of the dependent variable corresponding to neighboring cross-sectional units. The SAL model takes the form:

$$Y = \rho W_Y Y + X\beta + \varepsilon, \qquad \varepsilon \sim IID(0, \sigma^2 I), |\rho| < 1 \tag{2.4}$$

where Y is the dependent variable; $W_Y Y$ is a spatially lagged dependent variable; X is a matrix of explanatory variables; ε is a vector of independent and identically distributed (IID) homoskedatic error terms; and ρ is the spatial autoregressive parameter. Anselin (1993) refers to this model as a spatial autoregressive model with substantive spatial dependence. The presence of the spatial lag $(W_Y Y)$ as an explanatory variable in the model induces correlation with the error terms similar to an endogenous variable. In fact, the spatial lag for a particular observation $(W_Y Y)_i$ is correlated not only with the respective error term (ε_i), but also with the error terms of all other observations (ε_j). Estimating a non-spatial model – thus, ignoring the presence of the spatially lagged dependent variable – results in a misspecification error akin to an omitted variable (Anselin and Bera, 1998). Consequently, the OLS parameter estimates are biased and inconsistent. Some exceptions are discussed in Lee (2002)[12].

The model in the reduced form becomes:

$$Y = (I - \rho W_Y)^{-1} X\beta + e, \qquad e = (I - \rho W_Y)^{-1} \varepsilon \tag{2.5}$$

[12] The results in Lee (2002) are generated by a very special structure of the spatial weights matrix.

with the following covariance matrix: $V(e) = \sigma^2[(I - \rho W_Y)'(I - \rho W_Y)]^{-1}$. The covariance matrix indicates a global dependence structure; every observation is correlated with every other observation in the sample, but the strength of the interaction decays with distance (or order of contiguity), following Tobler's (1970) first law[13]. This relationship is made more explicit by expanding the spatial multiplier $(I - \rho W_y)^{-1}$ as follows[14]:

$$(I - \rho W_Y)^{-1}\varepsilon = \left(I + \rho W_Y + \rho^2 {W_Y}^2 + \rho^3 {W_Y}^3 + \cdots\right)\varepsilon \tag{2.6}$$

This is the reason that, despite the similarity of the SAL model with a time-series model, the properties of the OLS estimator in the former model do not extend to the latter. In a time-series model, the lagged dependent variable is uncorrelated with the error term in the absence of serial correlation in the errors and consistent estimates can be obtained by OLS[15] (Anselin, 1988a). This is not the case for the SAL model regardless of the properties of the error term. Anselin (1988a) attributes this lack of analogy in the properties of the OLS estimator to the two-dimensional and multi-directional nature of spatial dependence. Due to the endogeneity induced by the spatial lag variable $(W_Y Y)$, this model can be consistently estimated in a maximum likelihood framework or by means of instrumental variables.

The interpretation of the spatial autoregressive parameter in the SAL model depends on the source of spatial dependence. When spatial interaction is a feature of the economic behavior under study, a statistically significant estimate for the spatial autoregressive parameter (ρ) indicates true contagion or substantive spatial

[13] Tobler's first law: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

[14] $(I - \rho W_Y)^{-1}$ is known as a spatial multiplier because, as (2.6) shows, a change occurring in a particular location affects first-order neighbors, second-order neighbors, third-order neighbors, and so on.

[15] In this case the covariance matrix is triangular rather than a full matrix (Anselin and Bera, 1998).

dependence (Anselin and Bera, 1998). In other words, it measures the extent of spatial externalities or spatial spillovers. Examples in the applied economics literature include, among others, farmers' technology adoption behavior in Case (1992), state expenditures and tax setting behavior in Case et al. (1993), and strategic interaction among cities in the choice of growth controls in California in Brueckner (1998). If the spatial scale of the data does not match the spatial scale of the phenomenon under study, the SAL model can be used to filter out the spatial autocorrelation due to the scale mismatch[16]. In this case, the spatial autoregressive parameter is a nuisance parameter, but allows for proper interpretation of the significance of the other model parameters. The SAL model in this context has been used, for instance, to model urban housing and mortgage markets since they operate at different scales (e.g. Can, 1992; Can and Megbolugbe, 1997; Anselin and Can, 1996).

Spatial dependence in the error term, known as spatial error dependence, is typically modeled under the assumption that the error term follows a spatial autoregressive process[17]. The SAE model can be written as follows:

$$Y = X\beta + u, \quad u = \lambda W_u u + \varepsilon, \quad \varepsilon \sim IID(0, \sigma^2 I), |\lambda| < 1 \tag{2.7}$$

where Y is the dependent variable; X is a matrix of explanatory variables; $W_u u$ is the spatial error lag; ε is a vector of uncorrelated and homoskedastic error terms; and λ is the spatial autoregressive parameter. This model is appropriate when there is no theoretical anticipation of spatial interactions among economic agents and it is useful for

[16] The SAL model can be expressed as: $(I - \rho W_Y)Y = X\beta + \varepsilon$, where $(I - \rho W_Y)Y$ is a spatially filtered dependent variable (Anselin and Bera, 1998). For alternative approaches to spatial filtering see Getis (1995).

[17] Other specifications of the spatial autoregressive error process can be found in Cliff and Ord (1981), Haining (1988, 1990), Kelejian and Robinson (1993, 1995). These specifications, however, have seen limited applications.

correcting or filtering out the spatial autocorrelation present in spatial data[18]. The spatial autocorrelation may be a result of measurement error or unobserved variables that may spillover across spatial units but are otherwise not essential to the model (Anselin and Bera, 1998).

The model in the reduced form becomes:

$$Y = X\beta + e, \qquad e = (I - \lambda W_u)^{-1}\varepsilon \tag{2.8}$$

which results in the following covariance matrix: $V(e) = \sigma^2[(I - \lambda W_u)'(I - \lambda W_u)]^{-1}$. This covariance matrix highly resembles the one in the SAL model indicating a global dependence structure that decays with distance or order of contiguity. The similarity between the two covariance matrices complicates specification testing to properly discriminate between spatial error dependence and spatial lag dependence. We turn to this issue shortly. The estimated coefficients using OLS are in general unbiased but inefficient due to the non-spherical disturbances (u). These results are more in line with the properties of the OLS estimator in time-series models. More efficient estimates can be obtained by using robust methods that explicitly take into account the structure of spatial error dependence. In this model, the spatial autoregressive parameter is interpreted as a nuisance parameter (e.g. Benirschka and Binkley, 1994). Anselin (1981) shows that OLS estimation cannot produce a consistent estimate of the spatial autoregressive parameter (λ); thus, the SAE model is also estimated with alternative estimation frameworks such as maximum likelihood. The consequences of ignoring spatial error dependence are generally argued to be less severe than ignoring spatial

[18] The SAE model can be expressed as: $(I - \lambda W_u)Y = (I - \lambda W_u)X\beta + \varepsilon$, where $(I - \lambda W_u)Y$ and $(I - \lambda W_u)X$ are the spatially filtered dependent variable and spatially filtered independent variables, respectively, and ε is the uncorrelated error term (Anselin and Bera, 1998).

lag dependence, because while the former involves statistical considerations, the latter involves theoretical considerations (LeSage, 1997a).

A more general spatial process combines both types of spatial dependence. This model is known as a spatial autoregressive model with autoregressive disturbances (SAL-SAE). The model can be written as follows[19]:

$$Y = \rho W_Y Y + X\beta + u, \qquad u = \lambda W_u u + \varepsilon, \quad \varepsilon \sim IID(0, \sigma^2 I), |\rho, \lambda| < 1 \tag{2.9}$$

where Y is the dependent variable; X is the matrix of explanatory variables; $W_Y Y$ and $W_u u$ are the spatial lags; ε is a vector of uncorrelated and homoskedastic error terms; and ρ and λ are the spatial autoregressive parameters. The spatial weights matrices, W_Y and W_u, are generally assumed to be different to allow for two different spatial processes. If $W_Y = W_u = W$, complications arise in the identification of the spatial autoregressive parameters, ρ and λ. Anselin (1980) shows that, in this case, ρ and λ are identified only when a set of nonlinear constraints is strictly enforced.

The reduced form model can be written as follows:

$$Y = (I - \rho W_Y)^{-1} X\beta + e, \qquad e = (I - \rho W_Y)^{-1}(I - \lambda W_u)^{-1}\varepsilon \tag{2.10}$$

which generates a more complex covariance structure:

$$V(e) = \sigma^2\{[(I - \rho W_Y)(I - \lambda W_u)]'[(I - \rho W_Y)(I - \lambda W_u)]\}^{-1}$$

Empirical applications of the SAL-SAE model are found in Case (1987,1991,1992), Case et al. (1993), Besley and Case (1995), etc. Anselin and Bera (1998) point out that the need for higher order processes can be considered to be more a result of a poorly specified spatial weights matrix than due to a more complex spatial data generating process. For instance, if the spatial weights matrix for a spatial lag process does not

[19] A more general but highly complex spatial process is the spatial autoregressive moving-average (SARMA). Refer to Huang (1984) for a treatment of the SARMA process. Applications of this process can be found in Brandsma and Ketellapper (1979) and Blommestein (1983, 1985).

capture all the dependence in the data, it will result in spatial error autocorrelation. Florax and Rey (1995) investigate the effects of misspecified weights in spatial linear regression models.

The SAL and SAE models can be derived as special cases of the more general SAL-SAE specification by adding some restrictions to the parameters. These models make the standard taxonomy of spatial linear regression models. A richer taxonomy of spatial linear models is developed in Anselin (2003) based on the distinction between global and local dependence and the way in which it translates into the incorporation in a regression specification of a spatially lagged dependent variable, spatially lagged explanatory variables, and spatially lagged error terms. The relevance of any taxonomy, however, depends on its empirical applications.

Diagnostic Tests for Spatial Dependence

While a theoretical argument should suggest the nature of spatial dependence to be specified in a model, a number of diagnostic tests are also available in the literature. A general test for spatial error dependence is Moran I. This test, an extension of the test statistic in Moran (1950), was presented by Cliff and Ord (1972). The Moran I test statistic takes the form:

$$I = \frac{N}{S_0}\left(\frac{e'W_u e}{e'e}\right) \tag{2.11}$$

where e is a vector of OLS residuals; W_u is the spatial weights matrix; and S_0 is a standardization factor ($S_0 = \sum_i \sum_j w_{ij}$)[20]. The test was initially suggested by Moran as a simple test for correlation between nearest neighbors, which explains the similarity of

[20] For a row-standardized spatial weights matrix, $S_0 = N$, thus the Moran I test statistic simplifies to: $I = e'We/e'e$.

this test statistic with the Durbin-Watson statistic (Anselin and Bera, 1998). A standardized Moran's index follows an asymptotically standard normal distribution, thus the Moran I test is implemented based on an asymptotically normal standard z-value. The test is easy to implement because it only depends on OLS residuals. But, it has low power because it is a general test; it does not test against a specific alternative. A similar general test can be found in Kelejian and Robinson (1992). Kelejian and Prucha (2001) derive and examine the large sample distribution of Moran I type test statistics for the SAL-SAE model and a variety of limited dependent variable models.

An alternative to the general tests discussed above is one which states explicitly an alternative hypothesis based on the data generating process. For instance, given the spatial process: $u = \lambda W u + \varepsilon$, we can test for $H_0: \lambda = 0$ versus $H_a: \lambda \neq 0$. Any of the testing principles –Likelihood Ratio (LR), Wald, and Lagrange Multiplier (LM) – can be used. These testing principles result in tests that are asymptotically equivalent. In practice, the LM test is more appealing because its implementation requires only OLS estimation under the null hypothesis. The Wald and LR tests require maximum likelihood (ML) estimation under the alternative hypothesis and the functional forms of these tests are more complex. Interestingly, Monte Carlo evidence (e.g. Godfrey, 1981; Bera and McKenzie, 1986) suggest that the LM is not inferior to the LR or Wald test with respect to power, although it does not use the information in the alternative hypothesis. An LM test for spatial error autocorrelation was suggested by Burridge (1980). The test takes the form:

$$LM(error) = \frac{\left(\frac{e' W_u e}{\sigma^2}\right)^2}{tr\,[W_u{}' W_u + W_u^2]} \sim \chi^2(1) \tag{2.12}$$

where tr represents the trace operator; e is a vector of OLS residuals; W_u is the spatial weights matrix; and σ^2 is an estimate for the error variance ($\sigma^2 = e'e/n$). Under the null hypothesis of no spatial dependence this test statistic follows a $\chi^2(1)$ distribution. This test is designed to test a single specification, namely spatial error dependence, assuming this is the correct specification of the model. In the presence of spatial lag dependence ($\rho \neq 0$), this test is not valid even asymptotically (Anselin and Bera, 1998).

A similar LM test for testing spatial lag dependence was suggested in Anselin (1988a). Given the model specification:$Y = \rho W_Y Y + X\beta + \varepsilon$, the test is formulated as: $H_0: \rho = 0$ versus $H_a: \rho \neq 0$. In this case several tests based on the ML principle are available as well, but the LM test is more attractive because it only requires OLS residuals. The test statistic takes a more complex form:

$$LM(lag) = \left[\frac{e'W_Y Y}{\sigma^2}\right]^2 \left[\frac{\sigma^2}{(W_Y X\beta)' M_x (W_Y X\beta) + tr[(W_Y{}' + W_Y)W_Y]\sigma^2}\right] \sim \chi^2(1) \qquad (2.13)$$

where tr denotes the trace operator; $W_Y Y$ is the spatial lag; β is a vector of model parameters; and $M_x = I - X(X'X)^{-1}X$. This test statistic also follows a $\chi^2(1)$ distribution under the null hypothesis of no spatial dependence. Similar to the test for spatial error dependence this test is valid only if the assumed specification of the model is the correct specification. Anselin (1993) suggests that the type of spatial dependence can be diagnosed by comparing the LM test statistics in (2.12) and (2.13). The test with the most significant value tends to point to the correct type of spatial dependence.

We can also test for concurrent spatial lag and spatial error dependence. Anselin(1988b) suggested two approaches. The first approach involves estimating an SAL-SAE model and testing jointly for $\rho = \lambda = 0$. An LM test for this approach based on OLS residuals is developed in Anselin (1988b). A shortcoming of this test is that if the

null hypothesis is rejected, it is not possible to infer whether the model misspecification is a result of spatial lag dependence or spatial error dependence. Another approach is to test for spatial error dependence in the presence of spatial lag dependence (H_0: $\lambda = 0$ when $\rho \neq 0$) to determine whether the SAL model fully accounts for the spatial dependence, or alternatively, test for spatial lag dependence in the presence of spatial error dependence (H_0: $\rho = 0$ when $\lambda \neq 0$) to determine if the SAE model accounts for all spatial dependence. The LM tests proposed in Anselin (1988b) for the second approach require nonlinear optimization or numerical techniques. As an alternative, Anselin et al. (1996) developed computationally simpler diagnostic tests that utilize OLS residuals based on the modified LM test developed by Bera and Yoon (1993).

Estimation of Spatial Linear Regression Models

The review of the estimation methods for spatial linear regression model in this section focuses on the classical estimation framework. For Bayesian approaches to spatial models see, for example, Hepple (1995a, 1995b), LeSage (1997b), etc. LeSage and Pace (2009) provides a more recent review.

Maximum likelihood estimation

In maximum likelihood estimation, the probability of the joint distribution of all observations (i.e. the likelihood) is maximized with respect to the model parameters. The ML approach for models of spatial interaction received its first comprehensive treatment by Ord (1975)[21]. Under regularity conditions, the ML estimator for these

[21] ML estimation for spatial models originates in the work of Whittle (1954) and Mead (1967). In addition to Ord (1975), ML approaches for spatial autoregressive models and spatial moving average models have been developed by Cliff and Ord (1973), Hepple (1976), Hordijk and Paelinck (1976), Haining (1978), Anselin (1980), Doreian (1982), Cook and Pocock (1983), Blommestein (1985),etc.

models achieves the desirable properties of consistency, asymptotic efficiency, and asymptotic normality[22]. The regularity conditions are generally satisfied when the structure of spatial interaction, expressed by the spatial autoregressive parameter and the spatial weights matrix, is non-explosive (Anselin 1988a).

Consider the model specification for the general spatial process in (2.10). Following the notation in Anselin (1988a), let $A = I - \rho W_Y$ and $B = I - \lambda W_u$. The log-likelihood function for a vector of observations Y based on a joint standard normal distribution for the error terms ε is given by:

$$L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{1}{2}\right)\ln|\Omega| + \ln|B| + \ln|A| - \left(\frac{1}{2}\right)v'v \qquad (2.14)$$
$$v'v = (AY - X\beta)'B'\Omega^{-1}B(AY - X\beta)$$

where Ω is the error covariance matrix and $v'v$ is the sum of squares of the transformed error terms[23]. It becomes apparent from (2.14) that maximizing the log-likelihood function of this model is not equivalent to minimizing the sum of squared errors. The determinants of the spatial terms of the Jacobian, namely $|A|$ and $|B|$, prevent the OLS estimates from being the ML estimates[24]. In fact, the difference in the estimates given by the two estimators depends on the magnitude of these determinants and becomes greater for larger values of the spatial autoregressive parameters (ρ and λ).

[22] The regularity conditions consist of a non-degenerate and continuously differentiable log-likelihood, bounded partial derivatives, positive definite and non-singular covariance matrices, and finite quadratic forms. In addition, the number of model parameters should be fixed and independent of the number of observations to avoid an incidental parameters problem (Anselin,1988a).

[23] Since $\Omega = \sigma^2 I$, then $v'v = \frac{1}{\sigma^2}\varepsilon'\varepsilon$, where ε is the vector of uncorrelated and homoskedastic error terms in (2.10).

[24] The Jacobian is given by: $J = det(\partial\varepsilon/\partial y) = |B||A|$. To avoid an explosive spatial structure, the Jacobian needs to satisfy $|B| > 0, \text{and } |A| > 0$, which indirectly constrains the range of the spatial autoregressive parameters to guarantee a well-behaved log-likelihood function over the parameter space (see also footnote 10).

The first-order conditions for maximizing the log-likelihood function in (2.14) involve taking the partial derivatives of this function with respect to the model parameters[25]:

$$\begin{aligned} \partial L/\partial \beta &= v' \, (\Omega^{-1/2} BX) \\ \partial L/\partial \rho &= -tr \, A^{-1} \; W_Y + v' \Omega^{-1/2} B W_Y Y \\ \partial L/\partial \lambda &= -tr \, B^{-1} \, W_u + v' \Omega^{-1/2} W_u (AY - X\beta) \end{aligned} \tag{2.15}$$

The ML estimates are obtained by setting the first-order conditions equal to zero and solving for the parameter values. The system of the first-order conditions is nonlinear in parameters, thus its solution is obtained by numerical means. For the SAL and SAE model specifications, solving this system of equations is simpler because conditional upon the respective spatial autoregressive parameter, the first-order conditions for the non-spatial parameters have an analytical solution, which are then used to concentrate the log-likelihood function. The concentrated log-likelihood function is nonlinear only in the spatial parameter and it is maximized by numerical optimization. We now turn the attention to these models.

For the SAL model, the log-likelihood function can be derived by setting $B = I$ in (2.14). Hence:

$$L = -\left(\frac{N}{2}\right) \ln(2\pi) - \left(\frac{N}{2}\right) \ln(\sigma^2) + \ln|A| - \frac{(AY - X\beta)'(AY - X\beta)}{2\sigma^2} \tag{2.16}$$

The ML estimates of β and σ^2 are obtained by the first-order conditions as:

$$\begin{aligned} \hat{\beta}_{ML} &= (X'X)^{-1} X'(I - \rho W_Y) Y \\ \hat{\sigma}^2_{ML} &= \frac{(Y - \rho W_Y Y - X\beta_{ML})'(Y - \rho W_Y Y - X\beta_{ML})}{N} \end{aligned} \tag{2.17}$$

[25] For derivation details see Appendix 6.A in Anselin (1988a).

For known ρ, these ML estimates are equivalent to the OLS estimates from the regression of a spatially filtered dependent variable $(I - \rho W_Y)Y$ on the explanatory variables X (Anselin and Bera, 1998). A concentrated log-likelihood function is obtained by substituting (2.17) into the log-likelihood function (2.16), which is then maximized numerically to generate an ML estimate for ρ[26]. The formal conditions that lead to the desirable asymptotic properties of the ML estimator for the SAL model have been established by Bates and White (1985) and Heijmans and Magnus (1986a, 1986b, 1986c). More recently, Lee (2004) investigated the asymptotic distributions for the ML estimator and quasi-ML estimator for the SAL model and showed that these estimators have the usual asymptotic properties of ML.

The likelihood function for the SAE model can be written by setting $A = I$ in (2.14) as follows:

$$L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln|\sigma^2| + \ln|B| - \frac{(Y - X\beta)'B'B(Y - X\beta)}{2\sigma^2} \tag{2.18}$$

The spatial error autocorrelation in the SAE model can be considered a special case of the general parameterized non-spherical error terms and the model can be seen as a generalized least squares (GLS) type model. Thus, the ML estimation of the SAE model can be carried out as a special case of the general framework of the ML estimation of GLS models with unknown parameters in the covariance matrix developed by Magnus (1978)[27]. The log-likelihood function can be rewritten as:

[26] Refer to Anselin and Bera (1998) for expressions of the concentrated log-likelihood functions for both the SAL and SAE models and Anselin (1980) for more technical details.

[27] The estimation of this model by feasible generalized least squares (FGLS) was developed more recently by Kelejian and Prucha (1999).

$$L = -\frac{1}{2}\ln|\Omega(\lambda)| - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(Y - X\beta)'\Omega(\lambda)^{-1}(Y - X\beta)}{2\sigma^2} \tag{2.19}$$

$$\Omega(\lambda) = [(I - \lambda W_u)'(I - \lambda W_u)]^{-1}$$

The first-order conditions for maximizing this log-likelihood function result in the GLS estimates of β:

$$\hat{\beta}_{ML} = [X'\Omega(\lambda)^{-1}X]^{-1}X'\Omega(\lambda)^{-1}Y \tag{2.20}$$

Conditional on λ, the ML estimates of β are equivalent to the OLS estimates obtained by regressing a spatially filtered dependent variable $(I - \lambda W_u)Y$ on spatially filtered independent variables $(I - \lambda W_u)X$ (Anselin and Bera, 1998). A similar solution of the first-order conditions for σ^2 yields a ML estimate for σ^2. To obtain a consistent estimate of λ, the solutions of the first-order conditions for β and σ^2 are substituted into the log-likelihood function to concentrate the log-likelihood function. An alternative approach that involves an iterative solution of the first-order conditions is provided by Magnus (1978). The desirable asymptotic properties of the ML estimator in the presence of unknown parameters of the error variance are shown by Magnus (1978), Rothenberg (1984), and Andrews (1986).

In addition to the numerical procedures required for nonlinear optimization, ML estimation is complicated further by the determinants in the Jacobian that need to be repeatedly evaluated in searching for the ML estimates of the spatial autoregressive parameters. The determinants of the spatial terms in the Jacobian are determinants of NxN matrices, thus require numerically intensive procedures. While in some cases it may be feasible to evaluate these determinants directly (e.g. Pace and Barry, 1997), approximations to simplify their computation have also been proposed (e.g. Ord, 1975; Griffith 1990, 1992).

Ord (1975) derived a simplification for these determinants in terms of the eigenvalues of the respective spatial weights matrices as follows:

$$|I - \rho W| = \prod_{i=1}^{N} (1 - \rho \omega_i) \tag{2.21}$$

where ω_i are the eigenvalues of W[28]. The convenience of this simplification largely depends on the size of the sample, because the precision of the numerical procedure required to evaluate the eigenvalues deteriorates with the size of the spatial weights matrix. For instance, Kelejian and Prucha (1999) found using ISML routines that the eigenvalues of a row-standardized matrix could not be reliably calculated for dimensions of W greater than 400. With sparse matrix routines in MatLab, the eigenvalues for larger matrices have been computed but with unverified accuracy[29]. Griffith (1992) explored several simplifications of the Jacobian in SAL models for irregular lattices. In particular, a generalized Jacobian approximation seems to simplify substantially the computation of the Jacobian; however, its performance needs further examination[30].

In general, ML estimation of spatial autoregressive models is computationally intensive and requires specialized statistical software and advanced programming techniques. This is even more the case with empirical applications that involve micro-level spatial data sets with large sample sizes.

[28] The evaluation of the determinants using eigenvalues is simpler for symmetric matrices because the eigenvalues are real. However, it becomes challenging when the spatial weights matrix is not symmetric and when the sample size is large.

[29] See Bell and Bockstael (2000) for a discussion.

[30] The functional form of the generalized Jacobian approximation by Griffith (1992) is:
$J = 0.2\ln(1.75) + 0.12\ln(1.05) - 0.2\ln(1.75 + \rho) - 0.12\ln(1.05 - \rho)$

Instrumental variables/generalized method of moments estimation

The spatial lag $(W_Y Y)$ included as a regressor in the SAL model (2.4) suggests that the endogeneity (or simultaneity) that it induces in the model can be addressed using instrumental variables. The instrumental variable (IV) approach is based on the existence of an instrumental variable(s) that is strongly correlated with the endogenous variable $(W_Y Y)$, but asymptotically uncorrelated with the error term (ε). These conditions are formally expressed as[31]:

$$\begin{aligned} &E\ (Z'\varepsilon) = 0 \\ &E\ (Z'R) \neq 0, \quad R = [W_Y Y, X] \end{aligned} \tag{2.22}$$

where Z represents the set of instruments (including the exogenous variables X) and R is the matrix of explanatory variables. Consistent estimates for the model parameters $\theta = [\rho, \beta]$ are obtained by:

$$\hat{\theta}_{IV} = (Z'R)^{-1} Z'Y \tag{2.23}$$

provided that $Z'R$ satisfies the rank condition, hence it is invertible.

When there are more instruments available than endogenous variables in the model, the dimension of Z is greater than the dimension of R, thus the model is overidentified. In this case, the model can be estimated by using a generalized method of moments (GMM) framework[32]. The GMM framework provides a different interpretation for the IV estimator as a solution to the system of moment conditions:

$$Z'(Y - R\theta) = 0 \tag{2.24}$$

[31] Equivalently, these conditions can be expressed as: $plim\ \left(\frac{Z'\varepsilon}{N}\right) = 0$ and $plim\ \left(\frac{Z'R}{N}\right) = Q_{ZR}$, where Q_{ZR} is a finite and uniformly positive definite matrix.

[32] For a rigorous treatment of the GMM approach refer to Sargan (1958), Gallant and Jorgenson (1979), Hansen (1982), Bowden and Turkington (1984), White (1984), etc.

These moment conditions arise from a general optimization problem with the following objective function:

$$\min \theta(\theta) = (Y - R\theta)'ZMZ'(Y - R\theta) \tag{2.25}$$

where M is a positive definite matrix. When $M = (Z'Z)^{-1}$, the estimator of θ achieves minimum variance, thus it is efficient (Hansen, 1982). The solution to this optimization problem is the IV estimator:

$$\begin{aligned} \hat{\theta}_{IV-GMM} &= (R'P_Z R)^{-1} R'P_Z Y \\ P_z &= Z(Z'Z)^{-1}Z' \end{aligned} \tag{2.26}$$

where P_z is an idempotent projection matrix. Under exact identification, the IV-GMM estimator in (2.26) reduces to the IV estimator in (2.23). In addition, the former is computationally equivalent to the two-stage least squares (TSLS) estimator for simultaneous equations. The IV estimator has the desirable asymptotic properties of consistency and asymptotic normality, but it is, in general, not the most efficient estimator. The efficiency of this estimator largely depends on the choice of instruments (Anselin, 1988a).

The implementation of this estimation approach in the context of an SAL model seems straightforward, but a major implementation challenge is the proper choice of instruments for $W_Y Y$. A sensible instrument choice is $E[W_Y Y]$. This expectation, however, is a function of the unknown spatial parameter ρ:

$$E[W_Y Y] = W_Y E[Y] = W_Y (I - \rho W_Y)^{-1} X\beta$$

Kelejian and Prucha (1998) showed that for a row-standardized matrix W_Y and $|\rho| < 1$, $E[Y] = \left[\sum_{i=0}^{\infty} \rho^i \, W_Y^i\right] X\beta$, with ${W_Y}^0 = I$. Hence, the linearly independent columns in $[X, W_Y X, W_Y^2 X, W_Y^3 X, \ldots]$ can serve as instruments for $W_Y Y$. These instruments,

henceforth referred to as the KP instruments, have seen extensive application. In a similar vein, Lee (2003) argued an optimal instrument to be the expectation:

$$E[R] = E[X, W_Y Y] = [X, (I - \rho W_Y)^{-1} X\beta]$$

Some earlier choices of instruments based on spatially lagged predicted values, spatial lags of exogenous variables, or combinations of the exogenous variables have also been argued to give satisfactory results (e.g. Anselin, 1980, 1984)[33]. Having specified the set of instruments Z, we can obtain consistent estimates for this model by applying TSLS. In the first stage, $W_Y Y$ is regressed on Z to get the fitted values $\widehat{W_Y Y}$. In the second stage, the estimates of β and ρ are obtained by the regression of Y on R where $W_Y Y$ is replaced by $\widehat{W_Y Y}$. As a side remark, although this discussion focuses on the presence of the spatial lag variable $(W_Y Y)$ as the source of endogeneity, the IV approach can be easily applied to situations that involve additional endogenous explanatory variables.

The implementation of the IV/GMM approach for the SAE and SAL-SAE model specifications requires consistent estimates of the parameter(s) of the error covariance matrix. Despite the similarities between these models and the counterpart specifications in time-series analysis, the IV estimation procedures developed for time-series models have limited applications to spatial models[34]. Anselin (1988a) outlined an iterative IV approach and a nonlinear optimization approach for the SAL-SAE model but applications of these approaches have been almost non-existent and the properties of

[33] For instance, Anselin (1980) proposed the use of a spatially lagged predicted value of the dependent variable $W_Y \hat{Y}$, with $\hat{Y} = X(X'X)^{-1}X'Y$.

[34] The IV estimator was first advocated for time-series models (Anselin, 1988a). Recall that in time-series models a consistent estimate of the autoregressive parameter can be obtained by OLS as long as the errors are not serially correlated. This is not the case in the counterpart spatial specification.

these estimation procedures are largely unknown. The GMM approach to these spatial autoregressive models gained momentum by the work of Kelejian and Prucha in the late 1990s.

Kelejian and Prucha (1999) proposed a GMM estimator for the spatial autoregressive parameter (λ) in the SAE model and showed that the spatial parameter can be consistently estimated using FGLS based on the moments of the error terms. Recall the SAE model in (2.7). The model has two stochastic error terms u and ε that are related in the following way:

$$(I - \lambda W_u)u = \varepsilon, \qquad \varepsilon \sim IID(0, \sigma^2 I) \tag{2.27}$$

From the relationship of the error terms in (2.27), the three moments of interest are[35]:

$$\begin{aligned} E\left[\frac{1}{N}\varepsilon'\varepsilon\right] &= E\left[\frac{1}{N}u'(I - \lambda W_u)'(I - \lambda W_u)u\right] = \sigma^2 \\ E\left[\frac{1}{N}\varepsilon' W_u' W_u \varepsilon\right] &= E\left[\frac{1}{N}u'(I - \lambda W_u)' W_u' W_u (I - \lambda W_u)u\right] = \frac{\sigma^2}{N} tr(W_u' W_u) \\ E\left[\frac{1}{N}\varepsilon' W_u' \varepsilon\right] &= E\left[\frac{1}{N}u'(I - \lambda W_u)' W_u' (I - \lambda W_u)u\right] = 0 \end{aligned} \tag{2.28}$$

The moment conditions in (2.28) imply the following three-equation system:

$$\Gamma_N[\lambda, \lambda^2, \sigma^2] - \gamma_N = 0 \tag{2.29}$$

where Γ_N is a $3x3$ matrix containing functions of the error terms u; $\lambda, \lambda^2$and σ^2 are the parameters to be estimated; and γ_N is a $1x3$ vector that contains the remaining right-hand side expressions of the moment conditions. A sample analog of (2.29) in terms of the predicted values of error terms $(\hat{u})$ can be written as:

$$G_N[\lambda, \lambda^2, \sigma^2] - g_N = \nu_N(\lambda, \sigma^2) \tag{2.30}$$

[35] The last equality in (2.28) holds because the diagonal elements of the spatial weights matrix are zero by construction.

where G_N contains functions of predictors of u and $v_N(\rho, \sigma^2)$ can be thought of as a vector of residuals. Using OLS residuals as predictors of u and imposing the restriction between λ and λ^2, the generalized moments estimator for λ and σ^2 is the nonlinear least squares estimator corresponding to the following objective function:

$$\left(\hat{\lambda}_{NLS,N}, \hat{\sigma}^2_{NLS,N}\right) = argmin\,\{v_N(\lambda, \sigma^2)' v_N(\lambda, \sigma^2)\} \tag{2.31}$$

Once a consistent estimate of λ is obtained, the estimates for β are obtained by FGLS as follows:

$$\begin{aligned} \hat{\beta}_{FGLS} &= \left[X'\hat{\Omega}(\lambda)X\right]^{-1} X'\hat{\Omega}(\lambda)Y \\ \hat{\Omega}(\lambda) &= \hat{\sigma}^2_{NLS,N}\left[(I - \hat{\lambda}_{NLS,N}W_u)'(I - \hat{\lambda}_{NLS,N}W_u)\right]^{-1} \end{aligned} \tag{2.32}$$

This procedure is a straightforward application of nonlinear least squares and it does not involve computing the Jacobian determinants nor the eigenvalues of W_u. Thus, unlike ML estimation, it is feasible even for large samples. Bell and Bockstael (2000) provided the first application of this estimator to household-level data. They empirically assessed the performance of this approach relative to the conventional ML approach. Both estimators performed qualitatively similarly for sample sizes for which the ML estimation was feasible, leading the authors to conclude that this GMM approach is a good alternative to the ML procedures.

A generalized spatial two-stage least squares (GSTSLS) procedure for the SAL-SAE model specification was developed by Kelejian and Prucha (1998). The proposed three-stage procedure combines the GMM procedures for the SAL and SAE models and involves a Cochrane-Orcutt type transformation for the spatial model to correct for spatial autocorrelation. To outline this procedure, consider the SAL-SAE model in (2.9). The model can be written more compactly as:

$$Y = R\theta + u, u = \lambda W_u u + \varepsilon, \quad \varepsilon \sim IID(0, \sigma^2 I) \tag{2.33}$$

where $R = (W_Y Y, X)$ and $\theta = (\beta', \rho)$'.

To estimate this model, the first stage of the procedure consists of a TSLS estimation of the first equation of the model in (2.33) to get an estimate of θ and obtain the model residuals $\hat{u}$. TSLS is employed because the model cannot be consistently estimated by OLS since $E[(W_Y Y)u')] \neq 0$. The TSLS estimator is given by:

$$\hat{\theta}_{TSLS} = (\hat{R}'\hat{R})^{-1}\hat{R}'Y \tag{2.34}$$

where $\hat{R} = P_Z R = (X, \widehat{W_Y Y})$, $\widehat{W_Y Y} = P_Z W_Y Y$, $P_Z = Z(Z'Z)^{-1}Z'$, and Z is the set of KP instruments[36]. This TSLS estimator is consistent, but it does not utilize information relating to the spatial correlation of the error term.

In the second stage, the model residuals from the first stage, $\hat{u} = Y - R\hat{\theta}_{TSLS}$, are used to set up the error moment conditions in (2.28) and a consistent estimate of λ, is obtained by the GMM procedure for the SAE models developed in Kelejian and Prucha (1999)[37].

In the third stage, the model is re-estimated using TSLS after applying a Cochrane-Orcutt type transformation to account for the spatial autocorrelation. The transformed model takes the form:

$$Y_* = R_*\theta + \varepsilon \tag{2.35}$$

where $Y_* = Y - \lambda W_u Y$ and $R_* = R - \lambda W_u R$. The TSLS estimator of this model is then given by:

[36] The set of KP instruments in this case is extended to include subsets of the linearly independent columns of $(X, W_Y X, W_Y^2 X, \dots, W_u X, W_u W_Y X, W_u W_Y^2 X, \dots)$. See Kelejian and Prucha (1998) for more details.

[37] This estimation procedure was initially proposed in a working paper version of Kelejian and Prucha (1999) in 1995.

$$\hat{\theta}_{GSTSLS} = (\hat{R}_*'(\hat{\lambda})\hat{R}_*(\hat{\lambda}))^{-1}\hat{R}_*'(\hat{\lambda})Y_*(\hat{\lambda}) \tag{2.36}$$

where $\hat{R}_*(\hat{\lambda}) = R - \hat{\lambda}W_u R$, $Y_*(\hat{\lambda}) = Y - \hat{\lambda}W_u Y$, and $\hat{\lambda}$ is the consistent estimate of λ obtained from the second stage. Kelejian and Prucha (1998) also formally established the large sample properties of consistency and asymptotic normality of this estimator.

Both the IV/GMM and ML estimators of spatial autoregressive models have similar asymptotic properties. The main advantage of GMM estimation is that it is easy to implement irrespective of the sample size. The computational simplicity, in turn, allows for a larger set of problems to be analyzed spatially as well as it accommodates for more general and realistic spatial relationships (Bockstael and Bell, 2000). Moreover, the GMM estimator does not rely on normality of errors, thus it is consistent for departures from this assumption. When the errors are indeed normally distributed, the GMM estimator is less efficient than the corresponding ML estimator. Another potential disadvantage of the GMM framework is that it does not constrain the values of the spatial autoregressive parameters, as is the case in ML estimation, thus estimates of these parameters may lie outside the $[-1.1]$ interval.

As a final remark, when dealing with nonlinear spatial models such as spatial discrete-choice models, even the computationally simpler GMM estimation procedures become quite arduous. A discussion of the estimation challenges of discrete-choice models in the presence of spatial dependence is provided in the next chapter, but firstly a review of these models follows.

Economic Models of Discrete Choice

A discrete-choice model depicts choice behavior of economic agents who face a set of discrete economic alternatives. This probabilistic framework was introduced into

the economic literature by Marschak (1960) from the seminal work of Thurstone on psychophysical discrimination.[38] Marschak showed that the choice probabilities were consistent with utility maximization for utility functions with random components and introduced the Random Utility Maximization (RUM) model (McFadden, 2001). The RUM framework provides the theoretical justification for discrete-choice models as models of economic choice behavior (Quandt, 1956; McFadden, 1974). Applications of these models in the economic literature include studies of occupational choice, choice of transportation mode, firm and household location decisions, recreation demand, demand for differentiated products, etc.

Discrete-choice models belong to the larger class of qualitative response models[39]. A defining characteristic of these models is that the dependent variable is categorical, taking on a finite number of discrete outcomes. Based on the number of outcomes, categorical variables are classified as dichotomous (binary) and polychotomous (multinomial). Categorical variables with multiple outcomes can be further classified as unordered and ordered[40]. In unordered categorical variables the values assigned to each outcome can be arranged in any order since the order is not meaningful. An example of an unordered categorical variable is employment status – employed, unemployed, and out of the labor force. In contrast, in ordered categorical

[38] Thurstone (1927) modeled the response of human subjects to alternative levels of true stimuli as a series of pairwise comparisons based on perceived stimuli (perceived with some error). The interpretation of the perceived stimuli as satisfaction (or utility) led to a model of economic choice (McFadden, 2001).

[39] Qualitative response models also include count data models. A common aspect of discrete-choice models and count data models is that the dependent variable in both cases assumes discrete values. As a point of departure, the dependent variable in count data models although discrete is not categorical. As a result, count data models and discrete-choice modes require different estimation methods.

[40] Sequential is often listed as a third category. For our purpose, a sequential categorical variable can be thought of as an ordered variable with a particular ordering scheme; for example, education status (no high school/high school but no college/ college but no professional degree/ professional degree).

variables the outcomes are inherently ordered; hence, the values assigned to these outcomes are no longer arbitrary. An example of an ordered categorical variable is health status – poor, good, and excellent. In this case, the quantitative values assigned to the qualitative outcomes need to convey that an excellent health status is better than good, which, in turn, is better than poor. Excellent treatments of discrete-choice models in the literature can be found in Maddala (1983), Wooldridge (2002), Cameron and Trivedi (2005), and Train (2007), among other references. This review attends exclusively to multinomial choice models, and, in particular, to logit-based models.

Random Utility Maximization Framework

Unordered choice alternatives

Discrete-choice models are generally derived within a RUM framework assuming utility-maximizing behavior of economic agents[41]. An economic agent faces a decision that involves making a choice from a choice set that contains a finite number of mutually exclusive and exhaustive alternatives. Given the choice set, a rational agent with well-defined preferences is expected to choose the alternative that maximizes his utility subject to a (budget) constraint, or equivalently, the alternative that provides him the highest indirect utility[42]. The researcher, although observes the decision maker's choice, does not directly observe the decision maker's utility. Hence, it is further assumed that there are factors that jointly determine the agent's utility – thus, his choice – some of which are unobservable. From the researcher's viewpoint, the decision maker's utility is

[41] This discussion closely follows Train (2007) and Cameron and Trivedi (2005). Discrete-choice models are also consistent with other types of behavior (Train, 2007).

[42] Rational choice theory assumes that the agent's preferences are complete (all choices can be ranked in order of preference) and transitive (if X is preferred to Y, and Y is preferred to Z, then X is preferred to Z), which combined with a mutually exclusive and exhaustive choice set allows an individual to consistently rank all alternatives in terms of his preferences and make an unambiguous choice.

a random function. In this setting, the goal of the researcher is to understand the decision making process that led to a particular choice, having observed the choice and a number of choice- and individual-related factors (Train, 2007).

To set up the model, consider an individual (i) who makes a choice among L (unordered) alternatives available in the choice set C, where $C = \{1, \dots, L\}$. Define latent variables U_{il} to denote the indirect utilities associated with different choice alternatives. These utilities are known to the decision maker but are unknown to the researcher. Define an outcome variable Y_i to take value l if the lth alternative is chosen by the decision maker, $l \in C$. The decision maker's choice (Y_i) is observed by the researcher. The decision maker chooses the alternative that provides the greatest indirect utility (assuming no ties); hence he chooses a particular alternative from the choice set, say k, if and only if:

$$U_{ik} > U_{il}, \forall k \neq l, l \in C \tag{2.37}$$

Equivalently, the researcher observes:

$$Y_i = k \quad if \quad U_{ik} = \max_{\forall\, l \in C} \, (U_{i1}, \dots, U_{iL}) \tag{2.38}$$

Although the decision maker's utility is not directly observable, the researcher observes some attributes of the choice alternatives (S_{il}) as well as some attributes of the decision maker (X_i) and can specify a function that relates these observed factors to the decision maker's utility as follows[43]:

$$V_{il} = V(R_i, \theta), \forall l \in C \tag{2.39}$$

[43] A clarification on notation: throughout this manuscript, l refers generally to the alternatives in the choice set whereas k (and m used later) refers to a particular alternative in the choice set; S_{il} denotes alternative-varying variables; X_i denotes alternative-invariant variables; γ denotes alternative-invariant parameters; β_l denotes alternative-varying parameters; R_i denotes all observed factors, $R_i = [S_{il}, X_i]$, and θdenotes all model parameters associated with the observed factors, $\theta = [\gamma, \beta_l]$.

where R_i denotes all observed attributes, $R_i = [S_{il}, X_i]$, and θ denotes all the parameters associated with the observed factors that enter in the specification of V. V_{il} is called a representative utility and differs from U_{il} because of the unobserved factors. These unobserved factors capture any unobserved variations in tastes, unobserved attributes of alternatives or individuals, and/or errors in perception and optimization by the decision makers (Maddala, 1983). Denoting the unobserved factors by ε_{il}, the indirect utility for any of the alternatives in the choice set can be expressed as:

$$U_{il} = V_{il} + \varepsilon_{il}, \forall l \in C \tag{2.40}$$

The decomposition of the utility into a systematic component (V_{il}) and a stochastic component (ε_{il}) allows the researcher to analyze the choice of the decision maker probabilistically[44]. Note that a higher value of the representative utility in (2.40) for a particular alternative does not guarantee the choice of that alternative. The unobserved factors could be such that a different alternative may be sufficiently better and the latter may prevail.

The unobserved factors (ε_{il}) are treated as a random vector $\varepsilon = (\varepsilon_{i1}, \dots, \varepsilon_{iL})$ with joint density $f(\varepsilon)$. Having specified this density, the probability that decision maker i chooses a particular alternative k is given by:

$$P_{ik} = P(Y_i = k|R_i) = P(U_{ik} > U_{il}, \forall k \neq l|R_i) = P(V_{ik} + \varepsilon_{ik} > V_{il} + \varepsilon_{il}, \forall k = l|R_i)$$

Thus, the probability that decision maker i chooses alternative k is the expected value of the following indicator function over all possible values of unobserved factors[45]:

[44] In the absence of the stochastic component (i.e. if all the factors were known), the decision maker's choice can be predicted exactly.

[45] The indicator function I takes on value one for a combined value of differences in V and ε that induces the decision maker to choose alternative k (i.e. if the expression in parentheses is true) and zero otherwise. Also note that ties would have zero probability.

$$P_{ik} = \int_{\varepsilon} I(\varepsilon_{il} - \varepsilon_{ik} < V_{ik} - V_{il}, \forall k \neq l) f(\varepsilon) d\varepsilon = F_k(R_i, \theta) \tag{2.41}$$

where $f(\varepsilon) = f(\varepsilon_{i1}, \dots \varepsilon_{il})$. The functional form for F_k (or more generally F_l) should be specified in a manner that ensures that the choice probabilities are statistically valid; $P_{il} \in [0,1]$ and $\sum_{l=1}^{L} P_{il} = 1$. Different specifications of F are obtained from different assumptions about the joint distribution of the error terms and result in different specifications of discrete-choice models. The two most common specifications are the logit model and the probit model.

The logit model is derived under the assumption that the error terms are distributed IID type I extreme value. This distributional assumption facilitates the estimation of the logit model because the choice probabilities in (2.41) take a closed-form solution. A critical aspect of this assumption is the independence of the error terms for different alternatives, which implies that the unobserved factors are uncorrelated across alternatives. The latter gives rise to an important property of the logit model known as the independence of irrelevant alternatives (IIA), which implies the same degree of substitutability among alternatives irrespective of the composition of the choice set. This property of logit models will be discussed later in greater detail.

The probit model assumes that the unobserved factors are distributed jointly normal. This model avoids the independence of errors assumption and accommodates for different patterns of correlations; however, the choice probabilities do not have a closed-form solution. Thus, they have to be evaluated numerically through simulation. A major advantage of the probit model is the flexibility in handling correlation across alternatives (and over time). A drawback is the more complex estimation procedures and its reliance on the normal distribution, which in some situations may not be

appropriate. Given the similarity between the normal and the logistic distribution, in practice, both the logit and the probit models produce (after proper scaling) very similar results. The choice of a proper distribution is difficult to justify on theoretical grounds[46] (Greene, 2000).

In addition to the choice of F, multinomial models with unordered choice alternatives also differ from the type of explanatory variables included in the model depending on whether or not these variables vary across alternatives. Alternative-varying regressors take on different values by alternative (and possibly by individual). For example, the cost of transportation in a transportation choice model largely depends on the mode of transportation. In this case, F in (2.41) for any of the choice alternatives takes the form:

$$F_l(S_{il}, \gamma) = F_l(S_{i1}\gamma, \dots, S_{iL}\gamma)$$

where S_{il} denotes alternative-specific attributes and γ is a vector of alternative-invariant parameters. In McFadden's terminology, this is known as a conditional model and it is appropriate in cases where the decision maker's choice is made, at least in part, based on observed attributes of each alternative. In contrast, regressors may be individual-specific but alternative-invariant. For instance, in the transportation choice model, socio-economic characteristics of a decision maker are the same regardless of the transportation modes available. In this case, F in (2.41) takes the form:

$$F_l(X_i, \beta_l) = F_l(X_i\beta_1, \dots, X_i\beta_L)$$

where X_i denotes individual-specific attributes and β_l is a vector of alternative-varying parameters. This model, simply referred to as a multinomial model, is appropriate in

[46] The logistic distribution has heavier tails compared to the normal distribution. It closely resembles a t distribution with seven degrees of freedom (Greene, 2000).

cases where attributes of alternatives are either unimportant or data on these attributes is not available. For instance, in a model of occupation choice, the earnings of an individual in all occupations are not known. In this case, the decision maker's characteristics such as education and past experience are used to explain his occupation choice. Other applications, such as the transportation choice example above, may involve both types of variables. In this case, a richer model that combines both types of regressors is employed. This model is known as a mixed model. This distinction between the types of explanatory variables, and the resultant discrete-choice models, is important not only for their applications, but also for practical reasons. As it will become more evident when discussing logit-based models, it is possible to change the format of the variables from alternative-invariant to alternative-varying; hence, the mixed and multinomial models have a conditional model formulation. In this regard, the conditional model is more general.

Finally, equation (2.41) brings up two important points for the identification and estimation of discrete-choice models. First, the decision maker's choice is based on the difference in indirect utilities. This has implications in terms of the explanatory variables that can be included in the model and the interpretation of model parameters because they capture differences across alternatives. Second, the indirect utility (U_{il}) is only determined up to scale, so the scale of utility needs to be normalized. This is typically done by normalizing the variance of the error terms[47].

[47] As Train (2007) points out, the scale of utility and the variance of the error terms are linked by definition. Hence, normalizing the scale of utility is equivalent to normalizing the variance of the error terms. For instance**,** when utility is multiplied by λ, the variance of each ε_{il} changes by λ^2: $var(\lambda\varepsilon_{il}) = \lambda^2\, Var(\varepsilon_{il})$. See Train (2007) for a detailed discussion on identification of discrete-choice models.

Ordered choice alternatives

Ordered choice models are better conceptualized by considering a rather different decision making scenario. The decision now involves the decision maker evaluating an experience (event) based on the satisfaction (utility) acquired from the experience. The choice set contains alternatives that correspond to different levels of the decision maker's utility, but the alternatives are ordered such that a higher ranked alternate is associated with a higher level of utility. In this scenario, the decision maker is expected to choose an alternative from the choice set that best represents his overall satisfaction from the experience.

To set up a RUM model, consider an individual (i) who makes a choice among L (ordered) alternatives in the choice set C. Define a latent variable U_i to denote the utility associated with the experience. U_i has multiple levels associated with different choice alternatives in the choice set. As before, the researcher does not observe this utility index, but observes the choice made. Define an outcome variable Y_i to take value l if the decision maker chooses the lth alternative to represent his overall utility level associated with the experience. The researcher then observes:

$$Y_i = l \;\; if \;\; \alpha_{l-1} < U_i < \alpha_l \,, l \in C \tag{2.42}$$

where α_{l-1} and α_l are utility levels that define a utility interval such that if the utility perceived by the decision maker falls within this interval, it results in the choice of alternative l.

The utility (U_i) can be decomposed into a systematic component (V_i) that is expressed as a function of observed factors (R_i) and a stochastic component due to unobserved factors (ε_i) as follows:

$$U_i = V_i + \varepsilon_i = V(R_i, \theta) + \varepsilon_i \tag{2.43}$$

Then, the probability that individual i chooses any alternative l in the choice set is:

$$P_{il} = P(Y_i = l|R_i) = P(\ \alpha_{l-1} < U_i < \alpha_l, \forall l|R_i) = P(\alpha_{l-1} < V_i + \varepsilon_i < \alpha_l, \forall l|R_i)$$

$$= P(\alpha_{l-1} - V_i < \varepsilon_i < \alpha_l - V_i, \forall l) = F(\alpha_l - V_i) - F(\alpha_{l-1} - V_i)$$

Hence,

$$P_{il} = F[\alpha_l - V(R_i, \theta)] - F[\alpha_{l-1} - V(R_i, \theta)] \tag{2.44}$$

where F is the cumulative distribution function (CDF) of ε. Different distributional assumptions about the error term result in different specifications of ordered models. An ordered logit specification results from assuming that the error term is distributed logistic. Similarly, an ordered probit model is obtained by assuming a standard normal distribution for the error term. Either distributional assumptions result in choice probabilities that are statistically valid[48].

The threshold values α_{l-1} and α_l of the utility U_i are subjectively determined by the decision maker, thus are not known to the researcher. They are estimated along with the rest of the model parameters that enter the systematic component of utility. These parameters are a result of the ordered feature of the model. The choice probabilities in (2.44) indicate that ordered choice models face similar identification issues with unordered models; the absolute level of utility is unimportant and the scale of utility needs to be normalized. However, the estimation of ordered models is simpler because the choice probabilities in (2.44) do not involve multidimensional integrals. The next section provides a general discussion on the estimation of discrete-choice models.

[48] In addition to the logistic and normal distributions, other distributions used for ordered models are, for example, the log-log, negative log-log, and complementary log-log for skewed data (Angresti, 1996).

Estimation of Discrete-Choice Models

A common estimation method for discrete-choice models in applications that involve cross-sectional data is maximum likelihood. Consider the choice probabilities:

$$P_{il} = P(Y_i = l|R_i) = F(R_i, \theta), \forall l \in C$$

where F is the CDF of the unobserved factors; R represents the observed factors; and θ represents a vector of model parameters to be estimated. As emphasized previously, specification of F should ensure that the choice probabilities are not only statistically valid, but also consistent with maximization of a random utility function[49]. Normal and logistic distributions for the error terms produce choice probabilities that satisfy both conditions and result in different model specifications, namely multinomial logit/probit, ordered logit/probit, nested logit/probit, etc.

To set up the likelihood function, redefine the dependent variable in a binary form by introducing L binary variables as follows:

$$d_{il} = I(Y_i = l), \forall l \in C$$

For each observation Y_i only one of the terms d_{i1}, d_{i2},…, d_{iL} will be non-zero. Thus, the likelihood of observing Y_i is given by:

$$f(Y_i) = P_{i1}^{d_{i1}} \text{x}\, P_{i2}^{d_{i2}} \text{ x} \ldots \text{x} P_{iL}^{d_{iL}} = \prod_{l=1}^{L} P_{il}{}^{d_{il}} \tag{2.45}$$

For a sample of N independent observations, the likelihood function of the entire sample is the product of the individual likelihoods:

$$L_N(\theta) = \prod_{i=1}^{N} \prod_{l=i}^{L} P_{il}{}^{d_{il}} \tag{2.46}$$

[49] The RUM consistency conditions are established by Williams (1977), Daly and Zachary (1979), and McFadden (1981).

where θ is the vector of parameters that enter the systematic component of the utility. The multiplicative form of this likelihood function makes it more convenient to work with the log-likelihood function, which is given by:

$$L_N^*(\theta) = \ln\left[\prod_{i=1}^{N}\prod_{l=i}^{L} {P_{il}}^{d_{il}}\right] = \sum_{i=1}^{N}\sum_{l=1}^{L} d_{il}\ln(P_{il}) \tag{2.47}$$

The log-likelihood function is maximized at the point where:

$$\frac{\partial L_N^*(\theta)}{\partial\theta} = \sum_{i=1}^{N}\sum_{l=1}^{L}\frac{d_{il}}{P_{il}}\frac{\partial P_{il}}{\partial\theta} = 0 \tag{2.48}$$

Hence, the ML estimates of θ are obtained as a solution to the first-order conditions in (2.48). These equations are nonlinear in θ, because P_{il} is a nonlinear function of θ. This estimator will be consistent if the choice probabilities are correctly specified such that $E[d_{il}] = P_{il}$, in which case the expectation of (2.48) equals zero (Cameron and Trivedi, 2005). For logit models, the ML estimates are obtained by an iterative procedure that uses any of the familiar numerical optimization algorithms such as Newton-Raphson, method of scoring, etc. Probit models, in turn, are computationally costly even for a small choice set because the probit choice probabilities involve multidimensional integrals (Maddala, 1983; Amemiya, 1985). ML estimates for probit models are typically obtained by simulation methods[50]. For ordered models, the ML estimates are easily obtained for both the logit and probit specifications because the choice probabilities do not involve multidimensional integrals. The ML estimator for discrete-choice models has the usual asymptotic properties of consistency and asymptotically normality.

In the presence of endogeneity or correlation across the error terms for different observations, the discrete-choice models can be estimated using a GMM framework.

[50] Refer to Hajivassiliou and Ruud (1994) and Keane (1993) for reviews of simulation estimation.

Provided that the choice probabilities are correctly specified, the GMM estimator will be the solution to the following moment conditions:

$$\sum_{i=1}^{N}\sum_{l=1}^{L}(d_{il} - P_{il})Z_i = 0 \tag{2.49}$$

where Z is vector of instruments that does not depend on d_l. A choice for Z can be, for example, $Z_i = \partial P_{il}/\partial\theta$, which makes the moment conditions in (2.49) similar to the first-order conditions in (2.48). For correctly specified choice probabilities, the expectation of (2.49) equals zero. Hence, the GMM estimator is consistent, but its efficiency depends on the choice of Z. The moment conditions in (2.49) are nonlinear in θ so obtaining GMM estimates involves nonlinear estimation procedures as well.

Once the model is estimated, the interpretation of the model coefficients requires further computation to obtain marginal effects or elasticitites. The model coefficients do not have a direct marginal effect interpretation because of the nonlinearity of the model. For some model specifications, there is not even a direct correspondence between the coefficient sign and the sign of the respective marginal effect. Marginal effects measure the change in the choice probabilities for a ceteris paribus change in a regressor. They are typically computed for each individual as:

$$ME_i = \frac{\partial P_{il}}{\partial R_i} \tag{2.50}$$

An average marginal effect is obtained by averaging the individual marginal effects. Elasticities provide an interpretation of the model coefficients in terms of percentage changes as they measure the percentage change in the choice probabilities for a percentage change in the value of a regressor. They are obtained as follows:

$$E_i = \frac{\partial P_{il}}{\partial R_i}\frac{R_i}{P_{il}} \tag{2.51}$$

The effect of an attribute change on choice probabilities can also be computed by taking first differences in predicted choice probabilities obtained before and after the attribute change.

Logit-Based Models

Conditional, multinomial, and mixed logit models

Logit-based models with unordered choice alternatives are derived by assuming that the error terms ε_l in (2.41) are distributed IID type I extreme value (or Gumbel) with density:

$$f(\varepsilon_l) = e^{-\varepsilon_l} e^{-e^{\varepsilon_l}}, \qquad -\infty < \varepsilon_l < \infty$$

and CDF:

$$F(\varepsilon_l) = e^{-e^{\varepsilon_l}} \tag{2.52}$$

The type I extreme value distribution is asymmetrical (right-skewed) with mean 0.58 and variance $\frac{\pi^2}{6}$ (Cameron and Trivedi, 2005). This distributional assumption implicitly normalizes the scale of utility (Train, 2007). For any two type I extreme value random variables, ε_k and ε_l, their difference $(\varepsilon_k - \varepsilon_l)$ is a random variable distributed logistic with mean zero and variance $\frac{\pi^2}{3}$ (Johnson and Kotz, 1970). Hence, for a logit model, F in (2.41) is specified as the logistic CDF:

$$F(\varepsilon_{ik} - \varepsilon_{il}) = \frac{e^{\varepsilon_{ik} - \varepsilon_{il}}}{1 + e^{\varepsilon_{ik} - \varepsilon_{il}}} \tag{2.53}$$

The logistic distributional assumption about the error differences results in the following closed-form formula for the choice probabilities in (2.41)[51]:

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_l e^{V_{il}}} \tag{2.54}$$

[51] See Appendix A for a complete derivation of the logit choice probability formula from a RUM model set up. The logit probability formula can also be easily derived from a pure statistical perspective.

The functional form of the choice probabilities in (2.54) ensures well-behaved probabilities, because P_{ik} is inevitably between zero and one and the sum of choice probabilities for all alternatives equals one. In addition, these choice probabilities have also been shown to be consistent with RUM maximization[52]. This functional form of choice probabilities simplifies substantially the estimation of logit models because the log-likelihood function is globally concave in parameters θ (McFadden, 1974).

The simplest specification for the systematic component of indirect utility (V_{il}) is linear in parameters as follows:

$$V_{il} = R_i\theta = S_{il}\gamma + X_i\beta_l \tag{2.55}$$

where S_{il} represents the attributes of the lth alternative for individual i; X_i represents the attributes of the ith individual; and γ and β_l are vectors of parameters relating alternative attributes and individual attributes to the decision maker's utility, respectively. Given the specification of the representative utility V_{il} in (2.55), the probability that individual i chooses alternative k is given by:

$$P_{ik} = \frac{e^{(S_{ik}\gamma + X_i\beta_k)}}{\sum_l e^{(S_{il}\gamma + X_i\beta_l)}} \tag{2.56}$$

The model with choice probabilities (2.56) is known as a generalized or mixed logit (MXL) model because it incorporates both alternative-specific and individual-specific attributes.

[52] The logit formula was first derived by Luce (1959) from the IIA assumption. Marschak (1960) showed that given the IIA assumption, the model was consistent with utility maximization. Marley, cited in Luce and Suppes (1965), showed that an extreme value distribution for the unobserved portion of the utility leads to the logit formula. McFadden (1974) completed the analysis by showing the opposite; that the logit formula necessarily implies that unobserved utility is distributed extreme value. This chronology is obtained from Train (2007).

If only the subset of the alternative-varying explanatory variables in (2.55) is included in the model such that $\beta_l = 0$ in (2.56), the resulting model is referred to as a conditional logit (CL) model. In the CL model, the choice probabilities are specified as:

$$P_{ik} = \frac{e^{S_{ik}\gamma}}{\sum_l e^{S_{il}\gamma}} \tag{2.57}$$

The parameter vector γ does not vary by choice. Since $\sum_{l=1}^{L} P_{il} = 1$, an equivalent model can be specified by defining S_{il} in terms of deviations of regressors from the values of one of the alternatives l and setting the corresponding S_{il} equal to zero (Cameron and Trivedi, 2005).

In contrast, if only alternative-invariant regressors are included in the model, such that $\gamma = 0$ in (2.55), the choice probabilities take the form:

$$P_{ik} = \frac{e^{X_i\beta_k}}{\sum_l e^{X_i\beta_l}} \tag{2.58}$$

This model is known as a multinomial logit (MNL) model[53]. Since $\sum_{l=1}^{L} P_{il} = 1$, there are only $L - 1$ probabilities that can be determined independently. Thus, a normalization is needed for the identification of the model parameters. A common normalization sets one of the parameters β_l equal to zero. The interpretation of the model results is then made relative to the normalized alternative.

The distinction between the types of explanatory variables and the resultant logit model specifications is important in empirical applications. For instance, land-use research uses primarily location specific data whereas marketing research relies on category specific information. As a result, a MNL model is more appropriate for

[53] Strictly speaking, all the models reviewed in this section are multinomial logit models. This terminology is used in the literature with the understanding that this model accommodates only for individual-specific regressors.

analyzing landowners' land-use choices whereas a CL model is more appropriate for analyzing consumers' purchasing choices.

An interesting feature of these models is that the MNL and MXL models can be viewed as special cases of the CL model and can be expressed in a CL formulation. To see this, let X_i be a $(K\text{x}1)$ vector of individual attributes and corresponding parameter vector β_l. This vector can be expressed as a vector of alternative-varying (individual) characteristics by defining a $(KL\text{x}1)$ vector (X_{il}) with all zero elements except the elements corresponding to the lth block, which are set equal to X_i. In addition, we can define a common parameter vector β of dimension $(KL\text{x}1)$ with elements β_l. More explicitly,

$$X_{i1} = \begin{pmatrix} X_i \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \dots\ X_{il} = \begin{pmatrix} 0 \\ \vdots \\ X_i \\ \vdots \\ 0 \end{pmatrix}, \dots X_{iL} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ X_i \end{pmatrix} \text{ and } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{pmatrix} \tag{2.59}$$

Then, $X_i\beta_l = X_{il}\beta$. Hence, the choice probabilities can be equivalently expressed as:

$$P_{ik} = \frac{e^{X_i\beta_k}}{\sum_l e^{X_i\beta_l}} = \frac{e^{X_{ik}\beta}}{\sum_l e^{X_{il}\beta}} \tag{2.60}$$

The transformation of individual-specific variables to alternative-specific variables essentially involves expressing these variables as interactions with a series of alternative-specific dummy variables. Identification of the parameters in the MNL model requires that the parameters for one of the alternatives be normalized to zero. This normalization is handled in the CL formulation by omitting the variable interactions with the dummy variables corresponding to the alternative chosen for normalization. The expression of the MNL and MXL models as CL models is of practical importance because a computer program written for alternative-varying regressors can be used to

estimate all three model specifications since it is possible to change the format of the variables from alternative-invariant to alternative-varying[54].

Independence of irrelevant alternatives (IIA)

A key aspect of the distributional assumption that leads to the logit choice probabilities is the independence of the error terms across alternatives. The independence of the errors assumption implies that the unobserved portion of the utility for one alternative is unrelated to the unobserved portion of the utility for another alternative and results in the IIA property[55]. This assumption is likely to be violated in situations where two alternatives share common unobserved attributes or when unobserved individual characteristics influence how the observed individual and alternative attributes affect choice.

The IIA property is unattractive because it implies the same degree of substitutability between choice alternatives regardless of the composition of the choice set; hence, it results in a restrictive substitution pattern. The IIA property can be formally expressed using the odds ratio between two alternatives in the choice set (k and) as follows:

$$\frac{P_{ik}|C_1}{P_{im}|C_1} = \frac{P_{ik}|C_2}{P_{im}|C_2} = \frac{e^{V_{ik}}/\sum_l e^{V_{il}}}{e^{V_{im}}/\sum_l e^{V_{il}}} = \frac{e^{V_{ik}}}{e^{V_{im}}} = e^{V_{ik}-V_{im}} \tag{2.61}$$

where C_1 and C_2 denote two different choice sets that contain alternatives k and m; $P_{ik}|C_1$ and $P_{im}|C_1$ denote the probability of the ith individual choosing alternative k and m from the choice set C_1 (and similarly for choice set C_2). This odds ratio depends only

[54] This is the programming route taken in this dissertation. Refer to Cameron and Trivedi (2005) for more details.

[55] All three models, CL, MNL, and MXL exhibit the IIA property.

on alternatives k and m, thus the odds of choosing k over m are the same irrespective of the rest of the alternatives in the choice sets C_1 and C_2 and their attributes. A classical example in the literature is the red-bus-blue-bus problem due to McFadden (1974). In the famous example, we would expect that adding a different color bus (red) to the choice set consisting of a car and a (blue) bus would not change the choice probabilities of choosing between a bus and a car (assuming commuters are indifferent about the color). Instead, the logit model predicts a decrease in choice probabilities for both the (blue) bus and the car once the new (red) bus is introduced in the choice set. Thus, the model overstates the overall probability of choosing a bus.

Alternatively, consider improving an attribute of alternative k. We would expect the decision maker to substitute away from the other alternatives in favor of the alternative with the improved attribute. The effect of this change in the probabilities of other alternatives, for example alternative m, is given by the following cross-elasticity[56] :

$$E_{imk} = \frac{\partial P_{im}}{\partial S_{ik}} \frac{S_{ik}}{P_{im}} = -\frac{\partial V_{ik}}{\partial S_{ik}} S_{ik} P_{ik} \tag{2.62}$$

This cross-elasticity depends only on alternative k. This means that, while the improvement in the attribute of alternative k increases the choice probability of alternative k, it reduces the choice probabilities for alternative m and all other alternatives in the choice set by the exact same proportion. Thus, the reason that the ratio of the probabilities in (2.61) is constant is because the change of an attribute of a third alternative affects both choice alternatives k and m in the same proportion. This pattern of substitution is called proportionate shifting (Train, 2007).

[56] See Appendix A for derivations of elasticites and marginal effects.

Despite the restrictive model assumptions, the logit model accommodates for variations in taste as well as for dynamics of repeated choice as long as these variations are related to the observed factors. Any variation in unobserved factors enters the unobserved component of utility and violates the IID assumption of the error terms (Train, 2007). Interestingly, the restrictive IIA property is also considered to be an advantage of these models (Domencich and McFadden, 1996). Because the relative probabilities of alternatives within a subset are unaffected by the alternatives omitted, the IIA property allows the model to be estimated consistently with a subset of the choice alternatives or from data on binomial choices (Hausman and McFadden, 1984). This is particularly helpful when the choice set is relatively large or when the researcher is only interested in examining a subset of alternatives.

Various testing procedures are available in the literature to test whether or not the IIA assumption holds. One way is to re-estimate the model on a subset of alternatives. If IIA holds, parameter estimates for the subset should not be significantly different from the ones with the full set of choice alternatives. A test statistic for this type of test is found in Hausman and McFadden (1984). Another test involves estimating the model with cross-alternative variables and testing for the significance of the effect of these variables. Significant effects indicate a violation of the IIA assumption. A procedure along these lines is developed in McFadden (1987). In more flexible specifications (e.g. general extreme value (GEV) models), IIA can be tested by testing model restrictions, provided the more general model is a proper specification (Hausman and McFadden, 1984; McFadden, 1987; Train el al., 1989). However, when IIA fails, these tests do not provide much guidance as to the correct model specification. McFadden (1974)

advocates that the CL and MNL models be used only in cases when the alternatives can be viewed as distinct and weighted independently by the decision maker.

Several models that relax the IIA assumption have also been suggested[57]. A multinomial probit model is theoretically attractive because it accommodates for relatively unrestricted correlation patterns, but it has practical limitations due to the complexity of estimation[58]. Another approach to introduce some correlation into the model is to estimate a nested logit model[59]. In a nested logit model, the decision making process assumes a tree-like structure with several decision layers. Similar alternatives are grouped into nests and the unobserved factors are permitted to be correlated across nests, but uncorrelated within nests. While the IIA property does not hold for alternatives in different nests, it holds for alternatives within nests. A limitation of the nested logit model is that the decision making process might not fit an obvious nesting structure. A third possibility involves specifying a more general model, known as a random parameters logit model, which decomposes the unobserved factors into a part that contains all the correlation and another part that is distributed IID extreme value[60]. This model specification is quite general and can approximate any discrete-choice model.

---

[57] For overviews see Ben-Akiva et al. (1997) and Horowitz et al. (1994).

[58] Refer to McCulloch and Rossi (1994) and McCulloch, Polson, and Rossi (2000) for a general discussion.

[59] Nested logit models belong to the class of generalized extreme value (GEV) models. They are derived by assuming that the unobserved portion of utility for all alternatives is jointly distributed generalized extreme value. See McFadden (1984) for a detailed treatment of these models.

[60] See McFadden and Train (2000) for a general discussion on random parameters logit (or probit) models and their properties in approximating general choice patterns. This model is also known in the literature as a mixed logit model. This terminology is avoided here to prevent confusion with the mixed logit model that combines individual-specific and alternative-specific variables.

Ordered logit model

An ordered logit (OL) model is obtained by assuming that the error term in (2.43) follows a logistic distribution with CDF:

$$F(\varepsilon_i) = \frac{e^{\varepsilon_i}}{1 + e^{\varepsilon_i}} \tag{2.63}$$

Hence, the choice probabilities in (2.44) take the form:

$$P_{ik} = \frac{\mathrm{e}^{(\alpha_k - X_i\beta)}}{1 + e^{(\alpha_k - X_i\beta)}} - \frac{\mathrm{e}^{(\alpha_{k-1} - X_i\beta)}}{1 + \mathrm{e}^{(\alpha_{k-1} - X_i\beta)}} \tag{2.64}$$

where X denotes the explanatory variables; β denotes the parameters that enter the systematic component of utility; and α_{k-1} and α_k are the utility threshold parameters. Despite the similarity in the specification of the systematic component of utility with a MNL model, an OL model fits the same slopes across all outcomes (β does not vary by alternative). This is known as the parallel slopes assumption. It implies that the observed factors equally affect the likelihood of a person being in any of the different ordered categories. If the parallel slopes assumption is not expected to hold, the model should be estimated by MNL despite the ordered nature of the outcomes (Boroorah, 2002).

One way to test for the validity of the parallel slopes assumption is to estimate the model with MNL and OL and perform a LR type test using the likelihood values from both models[61] (Boroorah, 2002). If the parallel slopes assumption does not hold, treating unordered outcomes as ordered is likely to bias the estimates. In contrast, if ordered outcomes are treated as unordered, failing to impose the ordered structure may

[61] This test is only suggestive because it is not a LR test; the restricted model (OL model) is not nested within the unrestricted model (MNL model).

lead to loss of efficiency. In the view of this tradeoff, an ordered model should be used when there is an unequivocal ordering of outcomes.

The discussion on the models of discrete-choice in this chapter assumes that the decision makers make their choices independently. However, economic decisions of individuals are at times influenced by choices made by neighbors, peers, or other individuals at different locations that face similar decisions. Hence, modeling individual discrete-choice decisions may require accommodating for interdependence in decision making. Spatial dependence in discrete-choice models is the focus of the next chapter.

CHAPTER 3
AN ESTIMATION APPROACH TO DISCRETE-CHOICE MODELS WITH SPATIAL LAG DEPENDENCE INVOLVING LARGE SAMPLES

Spatial Dependence in Discrete-Choice Models

Spatial dependence in discrete-choice models has been essentially handled in the following ways. Standard probit or logit models have been employed to estimate a model with spatial data (e.g. Wallace, 1988; McMillen and McDonald, 1991; McMillen et al., 1992). Spatial dependence has been dealt with indirectly by using subsampling techniques to purge the data of spatial dependence (e.g. Carrion-Flores and Irwin, 2004; Nelson et al., 2001; Newburn et al., 2006) or by including spatially derived variables (e.g. Staal et al., 2002) and spatially lagged characteristics to capture some of the spatial effects (e.g. Nelson and Hellerstein, 1997). Lastly, spatial dependence has also been modeled directly (e.g. Case 1992; McMillen, 1992; LeSage, 2000; Beron and Vijverberg, 2003, 2004). The estimation of spatial discrete-choice models, however, remains challenging.

A spatial specification of a discrete-choice model differs from a standard (non-spatial) specification because the resultant spatially correlated covariance structure prohibits the expression of the multivariate distribution associated with the likelihood function as the product of univariate distributions; thus, the estimation of spatial discrete-choice models results in a multidimensional integration problem (Fleming, 2004). The complexity of the functional form makes the direct maximization of the discrete-choice likelihood function quite difficult in practice. For instance, Beron and Vijverberg (2000) reported that the estimation of a spatially autoregressive lag (SAL) specification of a probit model lasted several hours for a dataset with 49 observations. With larger datasets, the ML estimation of spatial discrete-choice models becomes

computationally infeasible. Moreover, the covariance structure implies that the model error terms are both autocorrelated and heteroskedastic, thus estimating a standard model leads to inconsistent parameter estimates[1] (McMillen, 1992). Several spatial discrete-choice estimators in the literature provide consistent estimates under spatially dependent data[2]. Applications of these estimators, however, are limited to empirical problems that involve binary choices and relatively small samples.

To simplify the estimation of the spatial model, some of the proposed spatial estimation procedures focus on spatial heteroskedasticity and ignore spatial autocorrelation. Consequently, they produce estimates that are consistent but not efficient. Case (1992) developed an estimation methodology for a heteroskedastic probit model with a spatially lagged dependent variable. The estimation methodology involves transforming the model to produce homoskedastic errors and estimate the variance normalized version of the model using maximum likelihood. This estimation procedure is made possible by the block diagonal structure assumed for the spatial dependence[3]. This approach can only be used in those empirical applications that can justify the special form of the spatial dependence structure. For instance, Rincke (2006) used this estimation methodology to study policy innovation in local jurisdictions assuming that

[1] The inconsistency of standard (probit or logit) models is a result of spatial heteroskedasticity. Spatial autocorrelation affects efficiency when full spatial information is not utilized. McMillen (1992) points out that the heteroskedasticity induced in the model is artificial because it arises from the location of spatial units relative to the boundary (not from differences in spatial units). Heteroskedasticity in spatial models is also examined by Haining (1988).

[2] A more rigorous discussion of these estimation techniques is provided in Fleming (2004).

[3] Case (1992) assumes that spatial dependence in her technology adoption application arises between farmers that reside within a district (but not across districts). She specifies a block diagonal contiguity weights matrix where the only non-zero elements correspond to a block of households that reside in a district. The block diagonal structure simplifies the estimation of the model because it produces an algebraic expression for $(I - \rho W)^{-1}$.

the dependence in the policy adoption decisions of school districts arises within school districts in the same local jurisdiction.

Pinkse and Slade (1998) proposed a GMM estimator based on the moment conditions derived from a heteroskedastic maximum likelihood function of a probit model with a spatial autoregressive error (SAE) structure. In the model, the heteroskedastic error variances are expressed as a function of the unknown spatial autoregressive parameter. As a result, the GMM estimation involves an iterative procedure that requires repeated evaluation of the covariance matrix for different values of the spatial parameter. This, in turn, involves inversion of NxN matrices with each iteration, which makes the implementation of the procedure computationally intensive. To avoid the matrix inversion, Klier and McMillen (2008) proposed a linearization of this spatial estimator in the context of a SAL logit model. Another application of Pinkse and Slade's estimator is found in Flores-Lagunes and Schiner (2010), which is used to obtain first step estimates for sample selection models with spatial error dependence.

McMillen (1992) also considered a spatial binary probit model with heteroskedatic errors. The form of heteroskedasticity in the model, however, is not derived directly from the spatial dependence structure. A functional form for the heteroskedasticity is assumed based on a spatial expansion method and the model is estimated by weighted nonlinear least squares. The implementation of this estimator is fairly easy even in large samples and produces consistent estimates provided the form of heteroskedasticity is correctly specified.

Since it is unclear how heteroskedasticity-consistent estimators perform in small to moderate sample sizes – as they may lead lo large variances – McMillen (1995a)

conducted a Monte Carlo investigation of the performance of a standard probit and a heteroskedastic probit when the dataset exhibits both heteroskedasticity and autocorrelation. To discern between the spatial effects, autocorrelation was specified as a function of the distance between two locations to ensure that it did not induce heteroskedasticity and heteroskedasticity was introduced separately in the model. The simulation results indicate that the standard probit is preferred to the heteroskedastic probit for small sample sizes and low degrees of heteroskedasticity. For larger sample sizes and higher degrees of heteroskedasticity the heteroskedastic probit outperformed the standard probit. Autocorrelation was found to have little to no effect in the results.

Other estimation procedures account for spatial heteroskedasticity by taking into account the full covariance structure of the model. These full information spatial estimators rely on the expectation-maximization (EM) algorithm, simulation methods, and Bayesian techniques to solve the multidimensional integration problem (e.g. McMillen, 1992; LeSage, 2000; Beron and Vijveberg, 2003, 2004). They attempt to preserve the estimation structure implied by maximum likelihood by simulating the choice probabilities or parameter probability distributions. The parameter estimates are obtained from the simulated distributions rather than from the direct maximization of the likelihood function.

McMillen (1992) extended the EM algorithm for spatial specifications of binary probit models. The EM algorithm consists of an iterative two step procedure based on the likelihood function of the continuous latent variable[4]. The E-step takes the

[4] The EM algorithm was introduced by Dempster et al. (1977) for time-series models. Amemiya (1985) and Ruud (1991) provide an overview of the method and applications. The algorithm has been proposed for use in spatial models (Flowerdew and Green, 1989) and discrete-choice models (Greene, 1990).

expectation of the likelihood function conditional on the observed discrete variable and an initial set of parameter values. In the M-step, the expectation of the likelihood function is maximized. The procedure is then repeated until convergence. An advantage of this estimator is that it avoids the direct evaluation of the multidimensional integral in the likelihood function of the spatial probit model. The procedure has essentially two drawbacks. First, it does not produce standard errors for parameter estimates. These precision measures have to be obtained from the N-dimensional dependence structure[5]. More importantly, the procedure requires repeated inversion of NxN matrices, thus it is computationally intensive in moderate to large samples. Computationally simpler alternatives are explored in McMillen (1995b).

LeSage (2000) adopted a Bayesian approach based on Gibbs sampling (Geman and Geman, 1984) to estimate a spatial probit model[6]. Gibbs sampling is a Markov Chain Monte Carlo (MCMC) technique that uses sampling from a set of conditional posterior distributions of model parameters to create a Markov chain that converges in the limit to the true posterior distribution of the model parameters. This approach has some advantages over the EM estimator in McMillen (1992). The standard errors of the parameters can be directly obtained from the parameters' conditional distributions. In addition, it accommodates for general forms of heteroskedasticity and more complex likelihood functions than a probit model. Due to its flexibility, the Bayesian framework has been adopted in several recent studies that incorporate spatial dependence in

[5] McMillen (1992) provides a covariance matrix by interpreting the probit estimator as a weighted nonlinear least squares estimator conditional on the spatial parameters.

[6] The Bayesian Gibbs sampler in LeSage (2000) is an extension to the Bayesian Gibbs sampling for non-spatial discrete-choice modes in Albert and Chib (1993) and spatial models with continuous dependent variables in LeSage (1997b).

multinomial and dynamic settings[7] (e.g. Kakamu and Wago, 2005; Chakir and Parent, 2009; Wang and Kockelman, 2009). Akin to the EM algorithm, this is an iterative procedure that requires recurring NxN matrix inversion.

Simulation methods have also been used to evaluate the dependent probit likelihood function. Beron and Vijverberg (2003, 2004) proposed a recursive importance sampling (RIS) procedure, which computes the likelihood function by simulating the parameter probability distributions[8]. A major advantage of this procedure is that, unlike the previous estimation procedures, it directly evaluates the likelihood function of the spatial probit model, so LR tests can be administered to test for model specification. In addition, standard errors for estimates can be obtained from the sampling distributions. A major drawback is the high computational costs associated with the simulator as made apparent in the Monte Carlo study conducted in Beron and Vijverberg (2003). Bolduc et al. (1997) compared the Gibbs sampler and the RIS simulator in the context of a multinomial probit model with an SAE structure. They concluded that while both approaches yield similar results, the Gibbs sampler is conceptually and computationally simpler.

An alternative to the ML type estimators is to interpret the spatial model as a weighted nonlinear version of a linear probability model with a general covariance matrix (e.g. Amemiya, 1985; Judge et al., 1985). These estimators are formulated as GMM estimators but can be viewed as weighted nonlinear forms of TSLS (for the SAL

[7] LeSage and Pace (2009) provide an extensive overview of the framework and applications in spatial discrete-choice models and other limited dependent variable models.

[8] The RIS simulator is a more general version of the more familiar GHK simulator (Geweke, 1989; Hajivassiliou, 1990; Keane, 1994). See Vijverberg (1997) and Beron and Vijverberg (2003) for a detailed treatment of the RIS simulator.

specification) and FGLS (for the SAE specification)[9]. Proposed spatial estimators for probit models with dependent observations (but homoskedastic errors) based on this approach are found in Avery et al., (1983) and Poirier and Ruud (1988). Fleming (2004) discusses in some detail the GMM methodology for SAL and SAE specifications of a binary choice model based on the GMM approach developed for spatial linear models in Kelejian and Prucha (1998,1999). The GMM approach avoids the multidimensional integral in the likelihood function, but the parameter estimates are typically obtained through an iterative procedure as in Pinkse and Slade (1998).

Klier and McMillen (2008) proposed a spatial GMM estimator for a SAL specification of a binary logit model based on the GMM estimator developed in Pinkse and Slade (1998). To avoid the inversion of large matrices required for estimation, they proposed to linearize the spatial model around a convenient point of parameter values for which the initial value of the spatial parameter (ρ) is set equal to zero. The linearization reduces the model estimation into two steps: a standard logit model followed by two-stage least squares. Klier and McMillen (2008) investigated the small sample properties of their linearized spatial estimator using Monte Carlo simulations. The results from the Monte Carlo study indicate that the linearized spatial estimator performs remarkably well in identifying the degree of spatial dependence induced in the simulated data (as measured by the spatial parameter ρ) for a sensible range of spatial parameter values. Because of its promising performance and ease of estimation, an extension of the Klier and McMillen (2008) methodology to polychotomous choice

[9] See Fleming (2004) for more details.

models would be valuable to analyze various economic decisions with multiple choice alternatives at a micro level. This chapter fills this gap in the context of logit models.

In the next section, it is shown that the key result that enables the linearization of the spatial model in the binary choice case in Klier and McMillen (2008) holds in the multinomial choice case and their binary choice "linearized logit" model is extended to a multinomial setting. Furthermore, the extension in this chapter covers various polychotomous logit models, all of which have different empirical applications. An important contribution of this research to the spatial modeling literature is that the proposed spatial multinomial choice estimators are computationally feasible even in very large samples.

Spatial Logit Estimators for Large Samples

The estimation methodology developed in Klier and McMillen (2008) for a binary logit model can be readily extended to a multinomial setting. First, consider a model with unordered choice alternatives. To set up the spatial discrete-choice model, consider an individual (i) who makes a choice among L alternatives. As point of departure from the RUM model discussed in the previous chapter, the decision maker's utility from a given alternative is now a function of a nearby agent's utility and other factors:

$$Y_{ik}{}^{*} = \rho \sum_{j=1}^{N} w_{ij} Y_{jk}{}^{*} + X_i \beta_k + S_{ik}\gamma + \varepsilon_{ik} \tag{3.1}$$

with,

$$Y_i = k \quad \text{if} \quad Y_{ik}^{*} > Y_{il}^{*}, \forall k \neq l, l = \{1, \dots, L\}$$

where $Y_k{}^{*}$ is a latent dependent variable representing the underlying utility from a particular alternative k; w_{ij} denotes the spatial weights relating neighboring observations i and j; X and S_k denote alternative-invariant and alternative-varying covariates, respectively; ε_k denotes the vector of errors for alternative k; and β_k, γ, and

ρ are the model parameters. The vectors of errors for all alternatives $\varepsilon_1, \dots, \varepsilon_L$ are assumed to be independent and identically distributed. The spatial weights matrix is typically row-standardized such that $\sum_{j=1}^{N} w_{ij} = 1$ for $j \neq i$, and $w_{ii} = 0$. The spatial autoregressive parameter (ρ) measures the degree of spatial dependence. A positive (negative) ρ implies that high values of the latent variable Y^* for observations neighboring observation i increase (decrease) the value of Y^* for observation i. The alternative with the highest latent utility is the one chosen by the decision maker, which is the choice observed in Y_i. The dependent variable Y_i can be defined in a binary form as: $d_{ik} = I(Y_i = k)$. The model in (3.1) forms the basis for a multinomial mixed logit (MXL) model with a spatially lagged dependent variable. Analogous SAL specifications for the conditional logit (CL) and multinomial logit (MNL) models are obtained as special cases of (3.1) with the proper subset of explanatory variables.

The reduced form of the spatial model in (3.1) is given by:

$$Y_{ik}^* = \sum_{j=1}^{N} \psi_{ij}\, X_j \beta_k + \sum_{j=1}^{N} \psi_{ij}\, S_{jk}\gamma + \sum_{j=1}^{N} \psi_{ij}\, \varepsilon_{jk} \tag{3.2}$$

with,

$$Y_i = k \quad \text{if} \quad Y_{ik}^* > Y_{il}^*, \forall l \neq k$$

where ψ_{ij} are the (i, j) elements of the spatial matrix $(I - \rho W)^{-1}$. The presence of the latter becomes more evident by writing the reduced model (3.2) in matrix notation as follows:

$$Y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} S\gamma + e, \qquad e = (I - \rho W)^{-1}\varepsilon \tag{3.3}$$

with the resultant covariance matrix: $V(e) \propto [(I - \rho W)'(I - \rho W)]^{-1}$. The model's covariance structure implies that the error terms (e) are both autocorrelated and heteroskedastic, which is the reason for the standard discrete-choice model to be

inconsistent. Following the notation in Klier and McMillen (2008), denote by σ_i^2 the variance of the error terms given by the diagonal elements of $V(e)$. A variance normalized version of the model is obtained by rescaling the model as:

$$\frac{Y_{ik}^*}{\sigma_i} = \frac{\sum_j \psi_{ij} X_j \beta_k}{\sigma_i} + \frac{\sum_j \psi_{ij} S_{jk} \gamma}{\sigma_i} + \frac{\sum_j \psi_{ij} \varepsilon_{jk}}{\sigma_i} \tag{3.4}$$

with, $$Y_i = k \quad \text{if} \quad \frac{Y_{ik}^*}{\sigma_i} > \frac{Y_{il}^*}{\sigma_i}, \forall k \neq l$$

The model in (3.4) implies the following choice probabilities:

$$P_{ik} = P(d_{ik} = 1) = P[\varepsilon_{il}^{**} - \varepsilon_{ik}^{**} < (X_i^{**}\beta_k + S_{ik}^{**}\gamma) - (X_i^{**}\beta_l + S_{il}^{**}\gamma)] \tag{3.5}$$

where $X^{**} = (I - \rho W)^{-1} X^*$, $X_i^* = X_i/\sigma_i$; $S^{**} = (I - \rho W)^{-1} S^*$, $S_{ik}^* = S_{ik}/\sigma_i$, $S_{il}^* = S_{il}/\sigma_i$; and $\varepsilon^{**} = (I - \rho W)^{-1} \varepsilon^*$, $\varepsilon_{ik}^* = \varepsilon_{ik}/\sigma_i$. If the error terms are assumed to be distributed IID type I extreme value, the difference of the error terms $(\varepsilon_{il}^{**} - \varepsilon_{ik}^{**})$ is distributed logistic. These distributional assumptions give rise to the spatial mixed logit (SMXL) model and the choice probabilities take the functional form:

$$P_{ik} = \frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}} \tag{3.6}$$

The present model can, in principle, be estimated using GMM (as in Pinkse and Slade, 1998) or nonlinear TSLS employing a set of instrumental variables Z and the model's gradients $\left(G = \frac{\partial P}{\partial \Gamma}\right)$. The spatial GMM estimator is the set of parameter values $\Gamma = (\beta_k, \gamma, \rho)'$ that minimizes the objective function $u'ZMZ'u$, where u denotes the vector of generalized model residuals; Z is a matrix of instruments; and M is a positive definite matrix (Z and M will be discussed later). The generalized model residuals are simply $u_{ik} = d_{ik} - P_{ik}$. If $M = (Z'Z)^{-1}$, the model reduces to nonlinear TSLS.

The gradients (G) for the SMXL model are[10]:

$$
\begin{aligned}
G_{\beta_{ik}} &= P_{ik}(\delta_{ilk} - P_{il})X_i^{**}, \quad \delta_{ilk} = I(l = k) \\
G_{\gamma_i} &= P_{ik}\left[S_{ik}^{**} - \sum_l P_{il}S_{il}^{**}\right] \\
G_{\rho i} &= P_{ik}\left[\left[\left(H_i\beta_k - \frac{X_i^{**}\beta_k}{\sigma_i^2}\Lambda_{ii}\right) + \left(K_{ik}\gamma - \frac{S_{ik}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right)\right]\right. \\
&\qquad \left. - \sum_l P_{il}\left[\left(H_i\beta_l - \frac{X_i^{**}\beta_l}{\sigma_i^2}\Lambda_{ii}\right) + \left(K_{il}\gamma - \frac{S_{il}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right)\right]\right]
\end{aligned}
\tag{3.7}
$$

where $H = (I - \rho W)^{-1}W(I - \rho W)^{-1}X^*$, $K = (I - \rho W)^{-1}W(I - \rho W)^{-1}S^*$, and $\Lambda = (I - \rho W)^{-1}W(I - \rho W)^{-1}(I - \rho W)^{-1}$. In this approach, the estimates are obtained through an iterative optimization process with the following parameter updating rule: $\Gamma_1 = \Gamma_0 + (\hat{G}'\hat{G})^{-1}\hat{G}'u_0$, where Γ_0 is an initial set of parameter values; $\hat{G}$ are the predicted values of the gradient terms from the regression of the gradients G on the set of instruments Z; and u_0 are the generalized model residuals. The algorithm converges when $\hat{G}'u_0$ approaches zero. Importantly, each step of this iterative procedure requires the inversion of the NxN matrix $(I - \rho W)$.

The main estimation insight from Klier and McMillen (2008) is to avoid the repeated inversion of $(I - \rho W)$ by linearizing the model — which in itself can be argued to be an approximation to the true model — around a particular point of initial parameter values. A convenient choice is $\Gamma_0 = (\beta_k^0, \gamma^0, \rho^0 = 0)'$, for which consistent estimates of β_k and γ are obtained from the standard MXL model. The linearization of the spatial model around Γ_0 is possible because, once we set the initial value of ρ equal to zero, the parameter ρ is still identified from the remaining terms H and K (both a function of ρ)

[10] The derivation of the gradient expressions for all the models considered in this section is relegated to Appendix B.

in the corresponding gradient in (3.7). At the linearization point, the gradients simplify greatly since $X_i^{**} = X_i^* = X_i,\ S_{ik}^{**} = S_{ik}^* = S_{ik}, S_{il}^{**} = S_{il}^* = S_{il}$; H and K become WX and WS, respectively; and the terms containing Λ_{ii} vanish because $\Lambda_{ii} = w_{ii}$, and $w_{ii} = 0$ by construction. Inversion of $(I - \rho W)$ is no longer needed as $(I - \rho W)^{-1} = I$. The gradient terms of the linearized SMXL model are:

$$
\begin{aligned}
G_{\beta_{ik}} &= P_{ik}(\delta_{ilk} - P_{il})X_i,\ \delta_{ilk} = I(l = k) \\
G_{\gamma_i} &= P_{ik}\left[\mathrm{S}_{ik} - \sum_l P_{il}S_{il}\right] \\
G_{\rho i} &= P_{ik}\left[[(WX)_i\beta_k + (WS)_{ik}\gamma] - \sum_l P_{il}[(WX)_i\beta_l + (WS)_{il}\gamma]\right]
\end{aligned}
\tag{3.8}
$$

The linearization of the spatial model is carried out by linearizing the generalized residuals u_{ik} around the initial parameter values Γ_0 as: $u_{ik} \approx u_{ik}^0 - G(\Gamma - \Gamma_0)$. The GMM estimator of the linearized model is the set of parameters $\Gamma = (\beta_k, \gamma, \rho)'$ that minimizes the objective function $v'Z(Z'Z)^{-1}Z'v$, where $v_{ik} = u_{ik}^0 + G\Gamma_0 - G\Gamma$. This estimation procedure is no longer iterative and involves two simple steps. First, the model is estimated with standard mixed logit, implicitly linearizing around the estimated mixed logit parameters and the initial value of ρ equal to zero. The initial parameter values $\hat{\Gamma}_0 = (\hat{\beta}_k^{MXL}, \hat{\gamma}^{MXL}, \rho^0 = 0)'$ are used to estimate the generalized model residuals $\hat{u}_{ik} = d_{ik} - \hat{P}_{ik}$ and the model gradients given by (3.8) $\hat{G} = (\hat{G}_{\beta_{ik}}, \hat{G}_{\gamma_i}, \hat{G}_{\rho_i})$. In the second step, each estimated gradient ($\hat{G}_{\beta_{ik}}, \hat{G}_{\gamma_i}$, and $\hat{G}_{\rho_i}$) is regressed on the set of instruments Z and fitted values ($\hat{\hat{G}}_{\beta_{ik}}, \hat{\hat{G}}_{\gamma_i}$, and $\hat{\hat{G}}_{\rho_i}$) are constructed. Finally, the coefficients in the regression of $(\hat{u}_{ik} + \hat{G}'\hat{\Gamma}_0)$ on $\hat{\hat{G}}_{\beta_{ik}}$, $\hat{\hat{G}}_{\gamma_i}$, and $\hat{\hat{G}}_{\rho_i}$ are the estimated parameter values of interest $\hat{\Gamma} = (\hat{\beta}_{ik}^{SMXL}, \hat{\gamma}^{SMXL}, \hat{\rho}^{SMXL})$.

Similar spatial specifications for the multinomial logit (MNL) and conditional logit (CL) models are obtained as special cases of (3.1) with the appropriate subset of explanatory variables[11]. A MNL model with a spatially lagged dependent variable is obtained by setting $\gamma = 0$ in (3.1) so the model allows only for alternative-varying covariates. The ensuing choice probabilities for the spatial multinomial logit (SMNL) model are:

$$P_{ik} = \frac{e^{(X_i^{**}\beta_k)}}{\sum_l e^{(X_i^{**}\beta_l)}} \tag{3.9}$$

Given these choice probabilities, the gradient terms for the SMNL model are:

$$\begin{aligned} G_{\beta_{ik}} &= P_{ik}(\delta_{ilk} - P_{il})X_i^{**}, \delta_{ilk} = I(l = k) \\ G_{\rho i} &= P_{ik}\left[\left(H_i\beta_k - \frac{X_i^{**}\beta_k}{\sigma_i^2}\Lambda_{ii}\right) - \sum_l P_{il}\left(H_i\beta_l - \frac{X_i^{**}\beta_l}{\sigma_i^2}\Lambda_{ii}\right)\right] \end{aligned} \tag{3.10}$$

where $H = (I - \rho W)^{-1}W(I - \rho W)^{-1}X^*$ and $\Lambda = (I - \rho W)^{-1}W(I - \rho W)^{-1}(I - \rho W)^{-1}$. The linearization of the SMNL model around $\Gamma_0 = (\beta_k^0, \rho^0 = 0)'$ is possible because the gradient of ρ in (3.10) is nonzero at Γ_0 due to the term H (a function of ρ), which does not vanish but reduces to WX. The gradients of the linerized SMNL model are:

$$\begin{aligned} G_{\beta_{ik}} &= P_{ik}(\delta_{ilk} - P_{il})X_i, \delta_{ilk} = I(l = k) \\ G_{\rho i} &= P_{ik}\left[(WX)_i\beta_k - \sum_l P_{il}(WX)_i\beta_l\right] \end{aligned} \tag{3.11}$$

Similarly, a SAL specification for a CL model is obtained by setting $\beta_k = 0$ in (3.1) to allow for individual-specific but alternative-invariant covariates. The choice probabilities for the spatial conditional logit (SCL) model are:

[11] Refer to Appendix B for the set up of the SMNL and SCL models and the corresponding gradient derivations.

$$P_{ik} = \frac{e^{(S_{ik}^{**}\gamma)}}{\sum_l e^{(S_{il}^{**}\gamma)}} \tag{3.12}$$

These choice probabilities result in the following model gradient terms:

$$\begin{aligned} G_{\gamma_i} &= P_{ik}\left[S_{ik}^{**} - \sum_l P_{il}S_{il}^{**}\right] \\ G_{\rho i} &= P_{ik}\left[\left(K_{ik}\gamma - \frac{S_{ik}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right) - \sum_l P_{il}\left(K_{il}\gamma - \frac{S_{il}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right)\right] \end{aligned} \tag{3.13}$$

where $K = (I - \rho W)^{-1}W(I - \rho W)^{-1}S^*$ and $\Lambda = (I - \rho W)^{-1}W(I - \rho W)^{-1}(I - \rho W)^{-1}$. The linearization of the SCL model around $\Gamma_0 = (\gamma^0, \rho^0 = 0)'$ is also possible since the spatial parameter ρ is identified from the remaining K terms (a function of ρ) in the corresponding gradient in (3.13), which simplify to WS. The gradients of the linearized SCL model are:

$$\begin{aligned} G_{\gamma_i} &= P_{ik}\left[S_{ik} - \sum_l P_{il}S_{il}\right] \\ G_{\rho i} &= P_{ik}\left[[(WS)_{ik}\gamma] - \sum_l P_{il}[(WS)_{il}\gamma]\right] \end{aligned} \tag{3.14}$$

The linearization of the SMNL and SCL models reduces their estimation to a two-step procedure analogous to the procedure described for the SMXL model with the corresponding initial set of parameters values and respective model gradients; $\hat{\Gamma}_0 = (\hat{\beta}_k^{MNL}, \rho^0 = 0)'$ and $\hat{G} = (\hat{G}_{\beta_{ik}}, \hat{G}_{\rho_i})$ for the SMNL model and $\hat{\Gamma}_0 = (\hat{\gamma}^{CL}, \rho^0 = 0)'$ and $\hat{G} = (\hat{G}_{\gamma_i}, \hat{G}_{\rho_i})$ for the SCL model. The parameter estimates from the two-step procedure are $\hat{\Gamma} = (\hat{\beta}_k^{SMNL}, \hat{\rho}^{SMNL})$ and $\hat{\Gamma} = (\hat{\gamma}^{SCL}, \hat{\rho}^{SCL})$, respectively.

For a SAL specification of a discrete-choice model with ordered choice alternatives, the decision maker's utility from a given alternative can be modeled as follows:

$$Y_i^* = \rho \sum_{j=1}^{N} w_{ij} Y_j^* + X_i\beta + \varepsilon_i \tag{3.15}$$

with,

$$Y_i = l \text{ if } \alpha_{l-1} < Y_i^* \leq \alpha_l, \forall l \in \{1, \dots, L\}$$

where Y^* denotes the latent dependent variable with observable counterpart Y; X denotes the explanatory variables; w_{ij} denotes spatial weights for observations i and j; ρ is the spatial lag parameter; β is a vector of model parameters; and ε is a vector of disturbances. The natural ordering of the alternatives is a result of the latent variable falling into various mutually exclusive and collectively exhaustive ranges given by the threshold parameters α_{l-1} and α_l, with $\alpha_0 = -\infty$ and $\alpha_L = \infty$.

The reduced from of the model is given by:

$$Y_i^* = \sum_{j=1}^{N} \psi_{ij} X_j \beta + \sum_{j=1}^{N} \psi_{ij}\, \varepsilon_j \tag{3.16}$$

with,

$$Y_i = l \; if \; \alpha_{l-1} < Y_i^* \leq \alpha_l, \forall l = \{1, \dots, L\}$$

where ψ_{ij} are the (i, j) elements of the inverse matrix $(I - \rho W)^{-1}$. In matrix notation, the reduced form can be written as:

$$Y^* = (I - \rho W)^{-1} X\beta + e, \quad e = (I - \rho W)^{-1}\varepsilon \tag{3.17}$$

with resulting covariance matrix: $V(e) \propto [(I - \rho W)'(I - \rho W)]^{-1}$. This covariance matrix also implies autocorrelated and heteroskedastic disturbances. As before, denote by σ_i^2 the variances of the errors given by the diagonal elements of the covariance matrix $V(e)$. Normalizing the model for heteroskedastic variances, we have:

$$\frac{Y_i^*}{\sigma_i} = \frac{\sum_j \psi_{ij}\, X_j \beta}{\sigma_i} + \frac{\sum_j \psi_{ij}\, \varepsilon_j}{\sigma_i} \tag{3.18}$$

with,

$$Y_i = l \text{ if } \frac{\alpha_{l-1}}{\sigma_i} < \frac{Y_i^*}{\sigma_i} \leq \frac{\alpha_l}{\sigma_i}, l \in C$$

Assuming that the model error term is distributed logistic, the choice probabilities for the spatial ordered logit (SOL) model are given by:

$$P_{il} = \frac{e^{(I_i^* \alpha_l - X_i^{**}\beta)}}{1 + e^{(I_i^* \alpha_l - X_i^{**}\beta)}} - \frac{e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}}{1 + e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}} \tag{3.19}$$

where $X^{**} = (I - \rho W)^{-1} X^*$, $X_i^* = X_i/\sigma_i$; $\varepsilon^{**} = (I - \rho W)^{-1}\varepsilon^*, \varepsilon_i^* = \varepsilon_i/\sigma_i$; and $I_i^* = 1/\sigma_i$ to make the scaling of the threshold parameters explicit. The choice probabilities in (3.19) result in the following model gradients:

$$\begin{aligned}
G_{\alpha_{l-1i}} &= -\left[\frac{e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right] I_i^* \\
G_{\alpha_{li}} &= \left[\frac{e^{(I_i^* \alpha_l - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_l - X_i^{**}\beta)}\right)^2}\right] I_i^* \\
G_{\beta_i} &= \left[\frac{e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^* \alpha_l - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_l - X_i^{**}\beta)}\right)^2}\right] X_i^{**} \\
G_{\rho i} &= \left[\frac{e^{(I_i^* \alpha_l - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_l - X_i^{**}\beta)}\right)^2}\right]\left[\left(-\frac{I_i^* \alpha_l}{\sigma_i^2}\Lambda_{ii}\right) - \left(H_i\beta - \frac{X_i^{**}\beta}{\sigma_i^2}\Lambda_{ii}\right)\right] \\
&\quad - \left[\frac{e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}}{\left(1 + e^{(I_i^* \alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right]\left[\left(-\frac{I_i^* \alpha_{l-1}}{\sigma_i^2}\Lambda_{ii}\right) - \left(H_i\beta - \frac{X_i^{**}\beta}{\sigma_i^2}\Lambda_{ii}\right)\right]
\end{aligned} \tag{3.20}$$

The linearization methodology proposed in Klier and McMillen (2008) to avoid the inversion of $(I - \rho W)$ can be used to estimate this SOL model only if the gradient with respect to the spatial lag parameter (ρ) in (3.20) is nonzero once we linearize the model around $\Gamma_0 = (\alpha_{l-1}^0, \alpha_l^0, \beta^0, \rho^0 = 0)'$. At the linearization point, $X_i^{**} = X_i^* = X_i$, $I_i^* = 1$, and all the terms containing Λ_{ii} become zero. The gradient of ρ, however, remains identified from H, which reduces to WX. Hence, the gradients for the linearized model are:

$$
\begin{aligned}
G_{\alpha_{l-1i}} &= -\left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2}\right] \\
G_{\alpha_{li}} &= \left[\frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i\beta)})^2}\right] \\
G_{\beta_i} &= \left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2} - \frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i\beta)})^2}\right] X_i \\
G_{\rho i} &= \left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2} - \frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i)})^2}\right] (WX)_i\beta
\end{aligned}
\tag{3.21}
$$

The estimation procedure for the linearized SOL model consists of the following two steps. In the first step, the model is estimated with standard OL to obtain the initial parameter values $\hat{\Gamma}_0 = \left(\hat{\alpha}_{l-1}^{OL}, \hat{\alpha}_l^{OL}, \hat{\beta}^{OL}, \rho^0 = 0\right)'$, which are used to compute the generalized residuals $\hat{u}_{ik} = d_{ik} - \hat{P}_{ik}$ and the model gradients in (3.21) $\hat{G} = (\hat{G}_{il-1}, \hat{G}_{il}, \hat{G}_{\beta_{ik}}, \hat{G}_{\rho_i})$. In the second step, each estimated gradient is regressed on the instruments Z and fitted values are obtained $(\hat{\hat{G}}_{il-1}, \hat{\hat{G}}_{il}, \hat{\hat{G}}_{\beta_{ik}}, \hat{\hat{G}}_{\rho_i})$. Finally, $\left(\hat{u}_{ik} + \hat{G}'\hat{\Gamma}_0\right)$ is regressed on $\hat{\hat{G}}_{il-1}$, $\hat{\hat{G}}_{il}$, $\hat{\hat{G}}_{\beta_{ik}}$, and $\hat{\hat{G}}_{\rho_i}$ to obtain the estimated parameter values of the SOL model $\hat{\Gamma} = (\hat{\alpha}_{l-1}^{SOL}, \hat{\alpha}_l^{SOL}, \hat{\beta}^{SOL}, \hat{\rho}^{SOL})$.

The linearization approach described in this chapter provides a good approximation to the underlying spatial models in (3.1) and (3.15) provided the latter are correctly specified and ρ is relatively small (i.e. closer to the linearization point). Applications of these spatial models are presented in subsequent chapters along with Monte Carlo evidence of the small sample properties of some of the proposed spatial estimators. Besides the empirical studies provided in this manuscript, another application of one of these spatial estimators, namely the SMNL estimator, can be found in Li (2010).

In order to carry out these estimation procedures, a remaining issue is the specification of the instrumental variables in Z. For a SAL model, it is common practice to specify Z to contain the linearly independent columns of $[X\ WX\ W^2X\ \ W^3X\ \ldots]$ as suggested by Kelejian and Prucha (1998). We employ the KP instruments for the Monte Carlo analysis and the two empirical studies that follow.

CHAPTER 4
SPATIAL ANALYSIS OF LAND-USE CONVERSION AT THE RURAL-URBAN FRINGE

Background

Urban decentralization and dispersion trends have led to increased conversion of rural lands in many urban peripheries and exurban regions of the United States (U.S). Moreover, growth in exurban areas is outpacing growth in urban and suburban areas, substantially changing the composition of the rural landscape. The growth pressure is particularly high at the rural-urban fringe[1].

Land-use changes occurring in the fringe and beyond warrant examination for the following reasons. First, these areas are under continuous land-use conversion, which is typically low-density and land-intensive. An analysis of the 1997 American Housing Survey Data conducted by the United States Department of Agriculture Economic Research Service found that, since 1994, 55% of the total land developed in the U.S. has been developed as housing lots greater than 10 acres and 90% as lots greater than 1 acre. In addition, 80% of all new development has occurred outside existing urban areas or in nonmetropolitan areas (Heimlich and Anderson, 2001). This form of land development – referred to as "urban sprawl" – is often argued to lead to inefficient and costly development patterns[2]. Second, direct connections exist between individual land-use conversion decisions and aggregate impacts of land-use changes (Bell and Irwin,

[1] The 1990 U.S. Census defines urban fringe as rural areas that are located in metropolitan counties. These areas are characterized by development that takes the form of new buildings (residential, commercial) and infrastructure, but does not meet the density requirement for urban classification. Exurban growth occurs in the rural areas beyond the urban fringe. It takes the form of large-lot development and consists of scattered single-family houses on large parcels of land.

[2] Although urbanization follows stages of growth that are well-understood, urban analysts debate whether "urban sprawl" is a natural outcome of well-functioning housing and land markets or a result of market failures. Settling this argument is beyond the scope of this work. Thus, when referring to urban development occurring at the fringe and beyond, we use the more neutral terms "urban development" or "urban growth".

2002). Thus, land-use conversion analysis at a micro-level (decision making level) provides valuable insights into the formation of land-use patterns in terms of timing of conversion and the spatial distribution of land uses. Development costs, such as local provision of public services, are shown to be a function of both the development patterns and the rate at which the conversion of land occurs (Frank, 1989; Ladd, 1992; Burchell et al., 1998). Lastly, individual landowners' decision making is driven by private incentives. Their decisions, however, may inflict economic, social, and environmental costs upon local communities that need to be addressed in public policy formulations. As the urban growth pressure keeps rising, imposing continuous restraints on open space and resources, it necessitates further examination of the land-use conversion process in order to obtain a better understanding of this process and the driving factors behind the land-use conversion decisions.

The economic literature on land use has been predominantly concerned with the efficient allocation of land as a scarce resource among competitive uses. One strand of this literature has examined the determinants and trends of land-use changes (e.g. Brown et al., 2005; Lubowski et al., 2008). Another strand has analyzed urban sprawl and its consequences (e.g. Ottensmann, 1977; Peiser, 1989; Ewing 1997; Burchell et al., 1998; Rusk, 1999; Brueckner, 2000). A third area has focused on the evaluation of existing land conservation programs (e.g., Kline and Alig, 1999; Newburn et al., 2006; Towe et al., 2008) and growth management controls (e.g. Fischel, 1990; Navaro and Carson, 1991; Feiltson, 1993; Nelson and Moore, 1996; Nelson, 1999; Kline, 2000). In the light of rising environmental concerns, more recent work has shifted focus towards quantifying the linkages between land-use changes and ecosystem changes with the

purpose of forecasting such changes (e.g. Lewis, 2010). While recognizing that land-use conversion is a spatial process associated with spatial externalities (and spillovers), this literature has only occasionally explicitly modeled the spatial dependence that characterizes the land-use conversion process. Because of the estimation complexity of the spatial models, spatial dependence has either been overlooked (e.g. McMillen, 1989) resulting in potentially inconsistent estimation, or spatially disperse data have been aggregated, resulting in artificially sharp intraregional differences or unrealistic interregional uniformity (Bockstael, 1996).

The purpose of this study is to examine the factors that drive landowners' land-use conversion decisions while investigating the possibility that their decisions are spatially interdependent. A landowner's decision to convert a given parcel of land to a particular land use is hypothesized to depend, among other factors, on the propensity of landowners of nearby parcels to choose that particular use. This suggests that the type of land conversion that is likely to occur in one area may influence (or may be influenced by) the type of land conversion in adjacent areas. Accounting for the potential presence of spatial interaction among individual landowners is important for two reasons. From an empirical perspective, estimating a standard model with spatially dependent observations will likely result in biased estimates and invalid statistical inference. From a policy perspective, the presence of spatial interactions among landowners who constitute the supply side of local land markets suggests that land parcels do not develop in seclusion. As such, uncoordinated local land use policies designed at a small scale such as subdivision regulation and zoning while attempting to

manage growth at a local level may fragment urban development and result in suboptimal land-use patterns at a regional level.

To model individual landowners' land-use conversion decisions, we borrow an economic model of land-use conversion from the optimal time to development literature. The base of the economic model is a landowner who is assumed to make an inter-temporal, profit maximizing choice regarding the conversion of a parcel of land to some alternative use. Since only a number of factors that affect the stream of returns and the cost of conversion are observable to the analyst, the net returns are modeled as having a systematic observable portion and a random unobservable portion. This treatment of the net returns allows for a latent variable formulation of the optimal land-use decision and allows for probabilistic statements about the landowner's land-use choice. Since the data on land use is typically categorical and the choice of land use is mutually exclusive, the theoretical model leads naturally to an empirical discrete-choice framework.

The empirical model used in this study is analogous to other spatially-explicit land-use change models estimated at the parcel level within a discrete-choice framework (e.g. Landis and Zhang, 1998; Irwin, 2002; Carrión-Flores and Irwin, 2004). Unlike the preceding analyses that only consider two land uses, our spatial analysis encompasses four land-use categories – agricultural, residential, industrial, and commercial – to accommodate for the possibility that land conversion to different urban uses may involve different development processes. The hypothesized spatial interaction among landowners regarding land conversion decisions is represented as a spatial lag. Thus, we estimate a multinomial logit model with a spatially lagged dependent variable using the linearization approach developed in the previous chapter. This novel estimation

method is feasible to use with parcel-level data, allowing us to model the individual landowners' land-use conversion decisions at the appropriate scale while exploiting the richness of parcel-level data. The latter enables addressing not only the spatial dependence on the land-use change decisions, but also the spatial heterogeneity of the landscape and land-use policies.

The remainder of this chapter is organized as follows. First, some of the external costs and benefits that make the urbanization phenomenon policy-worthy are reviewed. Next, a model of land-use conversion is developed to account for the assumed spatial interaction among landowners of neighboring parcels in the form of a spatial lag. The subsequent section transitions the economic model into an econometric model to be estimated using the SMNL estimator developed in the previous chapter. This is followed by a Monte Carlo study of the finite sample properties of the spatial estimator. A description of the data and the geographic area of study – Medina County, Ohio – along with a discussion of the empirical results succeed. Conclusions are drawn in the final section.

Impacts of Urban Growth

The outward expansion of urban areas into rural areas and the resultant sprawling development patterns are an outcome of the interplay between private decisions and public policies. A fundamental aspect of the dynamics of the American society is that as individuals and businesses achieve economic success, they relocate (Bier, 2001). Households' preference for larger dwellings located in less dense areas is driven by a preference for more spacious living arrangements, natural amenities such as scenic views and recreational opportunities, quality of public services, etc. Potential benefits from low density development include neighborhoods with lower crime rates, flexible

transportation, access to more open space, more privacy, a better sense of place and community, cheaper land and fewer taxes, better air quality, and a geographic separation of residences from commercial and industrial activities (Gordon and Richardson, 1997). Firms' relocation decisions, on the other hand, come as a response to economic opportunities such as agglomeration economies, better access to input and output markets, favoring public policies (e.g. lower taxes or less stringent environmental regulations), etc. Some governmental policies have also encouraged exurban development. Examples of these policies are the expansion of highways (e.g. Heavner, 2000), income tax subsidies for housing at a federal and state level, extension of public utilities and zoning at a local level (e.g. Moss, 1977; Anas et al., 1998; Pasha, 1996; Anas, 2001). In addition, the spatial distribution of urban growth is largely shaped by technology advancements that continue to lower the cost of transportation and communication, facilitating a trend of a more dispersed economy. Nelson (1992) summarizes the drivers of urban growth into four main categories: the continued deconcentration of employment and the rise of exurban industrialization, the latent antiurban and rural location preferences of U.S. households, technology advancements that make exurban living possible, and development policy biases favoring exurban development over compact development.

Sprawling urban development is associated with regional economic, environmental, and social impacts that are both positive and negative. From an economic perspective, while urban growth enhances economic activity in the areas in which it is occurring, it also decentralizes economic growth. Positive impacts on rural communities involve opportunities for off-farm employment, niche market access,

increased values of rural lands, etc. Ahearn et al., (1993) report that an average farm household earns more off-farm income than on farm. Moreover, off-farm employment opportunities have been crucial to the survival of many farms (Alig and Ahearn, 2006). However, urban development also results in increased land rental rates, localized declines in agricultural economies, and loss or fragmentation of agricultural lands. Importantly, prime agricultural land is continuously being lost to urbanization (Gardner, 1977). Heimlich and Anderson (2001) assert that low-density and fragmented development patterns do not threaten the national food and fiber production, but change the agricultural product and service mix and may result in reduced production of some high-value or specialty crops.

Potential benefits from low-density development like better housing choices, improved infrastructure, freedom from public transportation, diverse location options for businesses and households, involve higher costs of provision of public services, higher transportation costs from traveling over longer distances, etc. In addition, they incur social costs such as potential conflicts between farmers and homeowners, alteration of the rural character, compositional change in rural population, and increased separation of the urban poor. Porter (1997) shows that changes in community attributes, such as the mix of residents or loss of open space, are often perceived as costs by long-term residents of the community.

Land-use changes are also responsible for a growing number of environmental concerns. They are considered a major driver of changes occurring in natural ecosystems (Naiman and Turner, 2000; Armsworth et al., 2004). Land use and land cover changes are ranked second after climate change for their effects on the

functionality of these ecosystems (Grimm et al., 2008), the urbanized land contributing a major share (Alberti, 2005; Collins et al., 2000). The distribution of the population at a lower density results in loss of open space, increased traffic congestion and pollution, degradation of water quality, and fragmentation or loss of natural areas. Various ecological studies conducted on exurban development suggest significant impacts on biodiversity (Hansen et al., 2005). A long-term result of continuous land conversion could be declines in populations of desirable plant and wildlife species and increases in populations of opportunistic species (Maestas, 2007). Furthermore, regional changes in land cover and land use have been found to have cumulative effects on global climate change.

Many of the aforementioned impacts take the form of externalities and incur costs that are not internalized (Burchell, 1998). Thus, inefficiencies in land-use patterns emerge. In the absence of market imperfections, an individual landowner's optimal land-use choice coincides with the socially optimal land-use choice. However, in the presence of externalities social action is needed to manage social and environmental resources, calling for growth management and smart growth initiatives. Hence, state and local governments have adopted preservation of open space programs, conservation and retention programs, land-use planning policies, and growth controls. The role of these policies is to alter the private incentives so as to narrow the divergence between private and socially optimal outcomes and direct urban growth toward more sustainable areas. A better understanding of the land-use conversion process would help policymakers design more effective regional environmental, growth, and development policies.

Economic Model of Land-Use Conversion

To frame the landowners' land-use conversion decisions, we borrow an economic model of land-use conversion from the optimal time to development literature (e.g. Arnott and Lewis, 1979; Fujita, 1982; Capozza, Helsley, and Mills, 1986; Capozza and Helsley, 1989). Consider a risk-neutral price-taking economic agent (the landowner) that owns parcels of land of uniform quality[3]. The landowner chooses an optimal land use for a parcel such as to maximize the present discounted sum of the expected stream of future net land returns (net of conversion costs)[4]. The landowner builds his expectations of future net returns based on information about historic and current return values[5]. In addition, the net returns are a function of land attributes, which can be classified in three categories: biophysical characteristics, location (neighborhood) characteristics, and land-use regulations. For instance, land attributes associated with the agricultural land use may include land biophysical characteristics such as soil type, slope, fertility level, water-holding capacity, etc.; location characteristics such as proximity to market, transportation routes, public infrastructure, etc; and land-use regulations such as subdivision ordinances and zoning. In the case of residential or commercial land use, the expected returns may be a function of the distance to urban centers, employment, shopping sites, neighborhood amenities, public services, zoning, etc. Lastly, some

[3] The uniform-quality assumption allows land-use choices for a heterogeneous parcel to be treated as the sum of land-use choices on constituent uniform-quality parcels when returns and conversion costs are approximately linear in land quality (Lubowski et al, 2008).

[4] An efficient land market produces a gap between the price of land at the boundary of an urban area (minus conversion cost) and the value of agricultural land rent because of expected future rent increases due to development (Capozza and Helsley, 1989).

[5] Schatzki (2003) provides some empirical evidence in support of the assumption that land returns tend to hold over time (i.e. follow a random walk).

factors that affect the stream of returns and the cost of conversion are not observed by the analyst, so net returns are composed of a systematic observable portion and a random unobservable portion. This treatment of the net returns entails a latent variable formulation of the optimal land-use decision and allows for probabilistic statements about the land-use choice by the landowner. Since the data on land use is typically categorical and the choice of land use is mutually exclusive, the theoretical model leads naturally to an empirical discrete-choice framework. Note that the assumptions above have to be modified to model land-use conversion under a growth management scenario, since the decision of the landowner is conditioned upon the planner's objectives.

Let R_{ilt} denote the returns from a parcel of land i currently in use l at time t and A_{ilt} denote all observed attributes associated with parcel i. Let $A_{ilt} = [Q_{ilt}, K_{ilt}]$, where Q_{ilt} denotes the attributes that directly affect land returns and K_{ilt} denotes those parcel attributes that affect returns through conversion costs. Note that l represents the benchmark land use at time t; land could be at an undeveloped state with the potential to convert to a developed state, or it could be at a developed state with the potential to convert to an alternative developed state[6]. The land returns for parcel i in use l are given by:

$$R_{ilt} = R(A_{ilt}) + \varepsilon_{ilt} \tag{4.1}$$

[6] Conversion of land typically occurs in one direction, from an undeveloped state to a developed state, and generally from a rural land use (e.g. agricultural, forest) to an urban land use (e.g. residential, commercial, industrial). The reverse conversion from urban uses to other land uses is, in comparison, much more costly making the land-use conversion decisions potentially irreversible. The model presented here, however, does not impose any restrictions on the direction of land conversion.

where ε_{ilt} represents the unobserved parcel attributes. Let Y_{ilt} denote the net returns from a parcel of land i in use l at time t. Then, the net returns for parcel i in use l are obtained as:

$$Y_{ilt} = \{R(Q_{ilt}) - C(K_{ilt})\} + \varepsilon_{ilt} = Y(A_{ilt}) + \varepsilon_{ilt} \tag{4.2}$$

where C is the expected marginal cost of conversion. Under the model assumptions, the dynamic optimization problem simplifies to a one-period optimization problem with the decision rule to choose the land use with the highest expected net return. Hence, the landowner will choose to convert parcel i in use l to an alternative use k if and only if:

$$Y_{ik} > max_{l \neq k} Y_{il}, \quad \forall l = \{1, \dots, L\} \tag{4.3}$$

where $Y_{ik}$and Y_{il} are the expected net returns of parcel i in competing uses l and k. Thus, parcel i will be converted to use k if the expected net returns from use k exceed the expected returns from the current use l or any other alternative land use in the choice set. If the inequality does not hold for any of the uses in the choice set, parcel i will remain in the original state l.

The decomposition of the net returns into a systematic observable portion and a random unobservable portion entails a latent variable representation of the problem. Consider the "true" net returns for parcel i under current land use l:

$$Y_{il}^{*} = Y(A_{il}) + \varepsilon_{il} \tag{4.4}$$

where A is observed but not Y^{*}. Then, parcel i will be converted to use k if and only if:

$$Y_{ik}^{*} > Y_{il}^{*}, \forall k \neq l, l = \{1, \dots, L\} \tag{4.5}$$

This reformulation of the optimal land-use decision can be interpreted in the context of pressure for conversion and allows for probabilistic statements about the land-use choice by the landowner. The probability that the landowner chooses land use k over land use l is the probability that the unobserved factors when combined with the

observed factors are such that the "true" net returns under land use k are greater than under land use l, resulting in that particular outcome. Thus, the probability that a parcel i will be converted from use l to use k is given by:

$$P_{ik} = P(Y_{ik}^* > Y_{il}^*) = P[Y(A_{ik}) + \varepsilon_{ik} > Y(A_{il}) + \varepsilon_{il}] \tag{4.6}$$

In addition to the parcel attributes, it is plausible that the landowner's decision to convert a given parcel of land from one use to another will also depend on the propensity of landowners of nearby parcels to choose a particular use (either strategically or collaboratively). For example, a parcel in agricultural use may be less likely to convert to residential use if it is surrounded by parcels that are likely to remain in agricultural use. This feature of the land use decisions naturally leads to dependence over space. In this case, the landowner's net returns (for parcel i) from a given alternative are a function of the observed attributes A_{il} and unobserved attributes ε_{il}, but now A_{il} is comprised of own parcel attributes X_{il} and nearby parcel js attributes (owned by a different landowner) that determine its "true" net returns for the given alternative:

$$Y_{il}^* = Y\left(Y_{jl}^*, X_{il}\right) + \varepsilon_{il} \tag{4.7}$$

As before, parcel i will convert to use k if and only if:

$$Y_{ik}^* > Y_{il}^*, \forall k \neq l, l = \{1, \dots, L\} \tag{4.8}$$

The current model implies the following probabilities of choosing alternative k:

$$P_{ik} = P(Y_{ik}^* > Y_{il}^*) = P\left[Y\left(Y_{jk}^*, X_{ik}\right) + \varepsilon_{ik} > Y\left(Y_{jl}^*, X_{il}\right) + \varepsilon_{il}\right] \tag{4.9}$$

Equivalently, the probability that the landowner chooses land use k is the expected value of the following indicator function over all possible values of unobserved factors:

$$P_{ik} = \int I\left[\varepsilon_{ik} - \varepsilon_{il} < Y\left(Y_{jl}^*, X_{il}\right) - Y\left(Y_{jk}^*, X_{ik}\right)\right] f(\varepsilon_i) d(\varepsilon_i), \forall k \neq l \tag{4.10}$$

where $f(\varepsilon_i) = f(\varepsilon_{i1}, \dots, \varepsilon_{iL})$. Thus, the probability that a parcel of land will convert to an alternative land use depends, among other things, on the underlying factors determining the probabilities of neighboring parcels to convert (or remain) in that particular land use.

The model presented in this section is a simplified representation of the land-use conversion process and has been widely used to model landowners' behavior (e.g. McMillen, 1989; Bockstael and Bell, 1997; Landis and Zhang, 1998; Irwin and Bockstael, 2002). The model can be made dynamic by recognizing that the returns of a parcel of land in different uses – thus, its probability of conversion – may depend on the parcel's initial state, parcel changes over time, and changes in the surrounding landscape (Bockstael, 1996).

Econometric Model

The choice probabilities for the land-use conversion model in (4.10) can be empirically estimated once a functional form is assigned to the observed portion of the latent net returns and a distributional assumption is made for the unobserved portion. A common approach for dealing with spatially interdependent decisions is to estimate a spatial autoregressive lag model after quantifying the spatial interactions using a spatial weights matrix W. We specify a spatial autoregressive lag model in the context of a multinomial logit model since the observed parcel attributes are location specific. The spatial multinomial logit (SMNL) model takes the form:

$$Y_{ik}{}^{*} = \rho \sum_{j=1}^{N} w_{ij} Y_{jk}{}^{*} + X_i \beta_k + \varepsilon_{ik} \tag{4.11}$$

with,

$$d_{ik} = I(Y_{ik}^{*} > Y_{il}^{*}), \forall k \neq l$$

where $Y_{ik}{}^{*}$ ($Y_{il}{}^{*}$) represents the latent land net returns of parcel i from land use k (l); w_{ij} denotes the spatial weights relating parcels in locations i and j; X_i denotes observed

parcel i's attributes; d_{ik} is a binary variable that takes a value of 1 if alternative k is chosen by the landowner of parcel i and zero otherwise; ε_{ik} represents unobserved parcel i's attributes associated with land use k; and ρ and β_k are the parameters of interest. For a multinomial logit model, the disturbances ε_k are assumed to follow an IID type I extreme value distribution. The spatial weights matrix is row-standardized such that $\sum_{j=1}^{N} w_{ij} = 1, \forall j \neq i$, and $w_{ii} = 0$. The alternative with highest latent net returns Y^* is the one chosen by the landowner, which is the land-use choice observed and indicated by d in (4.11).

The parameter ρ measures the strength of the spatial interaction between landowners of adjacent land parcels regarding their land-use change decisions. In this context, a positive ρ means that a high propensity of a given landowner to convert his land parcel to a particular land use positively affects the propensity of a landowner of a nearby parcel to convert the latter to the land same use, thus increasing the probability of conversion of both parcels to the given land use. A realization of the landowners' land conversion decisions under positive spatial dependence results in clusters of land parcels in similar use. Moreover, a higher ρ implies a stronger mutual influence of landowners of adjacent land parcels in their respective land conversion decisions.

A spatially lagged dependent variable has been incorporated successfully in models where the dependent variable is continuous (e.g. Case et al., 1993; Brueckner, 1998; Brett and Pinkse, 2000; Saavedra, 2000; Fredriksson and Millimet, 2002). The estimation of a spatially autoregressive lag model, however, is challenging in a discrete-choice framework. To estimate the model, we use the linearization methodology developed in the previous chapter which reduces the model estimation to two simple

steps: a standard multinomial logit model with no spatial dependence followed by TSLS estimation of the linearized model which accounts for the spatial dependence. More specifically, the estimation procedure consists of:

Step 1. The model is estimated with standard multinomial logit to obtain initial parameter values for β_k and the initial value of ρ is set equal to zero. The initial parameter values $\hat{\Gamma}_0 = (\hat{\beta}_k^{MNL}, \rho^0 = 0)$ are then used to compute the generalized model residuals $\hat{u}_{ik} = d_{ik} - \hat{P}_{ik}$ and model gradients $\hat{G} = (\hat{G}_{\beta_{ik}}, \hat{G}_{\rho_i})$. At the linearization point $\hat{\Gamma}_0$, the gradient terms for the SMNL model are given by:

$$\hat{G}_{\beta_{ik}} = \hat{P}_{ik}(\delta_{ilk} - \hat{P}_{il})X_i, \delta_{ilk} = I(l = k)$$
$$\hat{G}_{\rho i} = \hat{P}_{ik}\left[(WX)_i\hat{\beta}_k - \sum_l \hat{P}_{il}(WX)_i\hat{\beta}_l\right] \tag{4.12}$$

Step 2. In the first stage of the TSLS procedure, each estimated gradient in $\hat{G}$ is regressed on the set of instruments Z (KP instruments) and fitted values are constructed ($\hat{\hat{G}}_{\beta_{ik}}$and $\hat{\hat{G}}_{\rho_i}$). In the second stage, the "linearized" generalized residuals are computed as ($\hat{u}_{ik} + \hat{G}'\hat{\Gamma}_0$), which are then regressed on $\hat{\hat{G}}_{\beta_{ik}}$and $\hat{\hat{G}}_{\rho_i}$. The coefficients obtained from the last regression are the estimated parameter values of interest $\hat{\Gamma} = (\hat{\beta}_k^{SMNL}, \hat{\rho}^{SMNL})$.

Monte Carlo Analysis

Prior to estimating the land-use conversion model, we conduct a Monte Carlo study in order to assess the finite sample performance of the SMNL estimator[7]. The experimental design follows Klier and McMillen's (2008) design with the necessary adjustments for the multinomial choice extension. We consider four choice alternatives

[7] Similar simulation studies were also conducted to examine the finite sample properties of the SCL and SMXL estimators. Those studies generated qualitatively similar results.

– to make the simulations comparable to the land-use application – and examine two scenarios. In the first scenario, all four alternatives have the same choice probability of 0.25. To generate this data we start with the reduced form of the model (4.11) written in matrix notation as:

$$Y^* = (I - \rho W)^{-1} X\beta + e, \qquad e = (I - \rho W)^{-1}\varepsilon \tag{4.13}$$

with covariance matrix $V(e) = [(I - \rho W)'(I - \rho W)]^{-1}$. We generate a single explanatory variable (X) uniformly distributed in the interval $(-1, 1)$. Next, we specify W (see below) and set the value of ρ between 0 and 0.9, varying it in increments of 0.1. Having specified W and a value for ρ, the explanatory variable is first transformed to obtain X^* as $X_i^* = X_i/\sigma_i$, where σ_i is the square root of the error variances given by the diagonal terms of $V(e)$. A second transformation follows to obtain X^{**} as $X^{**} = (I - \rho W)^{-1} X^*$. For simplicity, each of the parameters β_k is set equal to 1. Equal parameter values for the different choice alternatives and a fixed X (thus, a fixed X^{**}) ensure equal choice probabilities. Subsequently, the simulated probabilities are obtained as:

$$P_{ik} = \frac{\mathrm{e}^{X_i^{**}\beta_k}}{\sum_l \mathrm{e}^{X_i^{**}\beta_l}} \tag{4.14}$$

To generate the observed individual choices based on the simulated probabilities, we generate a uniform $(0, 1)$ random variable (u) and set $d_{ik} = 1$ for $k = l$ if:

$$\sum_{k=0}^{l-1} P_{ik} < u < \sum_{k=0}^{l} P_{ik}, \text{ with } P_{i0} = 0 \tag{4.15}$$

In the second scenario, we modify the way we generate the data to accommodate for different choice probabilities. This is the case that is more likely to represent a real data situation. To generate this data, we first fix the choice probabilities and solve for the parameters β_k that correspond to the desired probabilities (Appendix C). Specifically, we set $P_0 = 0.10,\ P_1 = 0.50,\ P_2 = 0.25,\ P_3 = 0.15$ to resemble to some

extent to the proportion of land parcels in each land-use category in our dataset. The explanatory variable (X) is generated uniformly distributed in the interval $(0.5, 1.5)$ and transformed as described above to obtain X^* and X^{**}. The parameters β_k are now a function of the expected value of X^{**} – this is the reason for changing the interval of X to avoid a mean of zero – and as a result a function of ρ, hence denoted by $\beta_k(\rho)$. Having obtained X^{**} and the parameters $\beta_k(\rho)$, the simulated probabilities are generated as in (4.14) and the observed individual choices are assigned based on the simulated probabilities using the assignment rule in (4.15).

The spatial weights matrix W in both cases is specified as a row-standardized first-order contiguity matrix, setting the non-zero elements to: $w_{ij} = 0.5\ if\ |i - j| = 1$ with endpoints $w_{1,2} = w_{N,N-1} = 1$. This specification preserves the spatial structure of the model but simplifies considerably the model estimation. The matrix of instruments Z is specified to contain the linearly independent vectors in $[X\ WX\ W^2X\ W^3X]$. A naïve estimation procedure – a standard multinomial logit model which ignores the spatial effects (MNL) – is used as an estimation benchmark. To summarize the simulation results, we report the average bias and root mean-squared error (RMSE) of the three estimated slopes (one is normalized to zero) and the spatial parameter estimated by the SMNL model. In general, we expect that if the true spatial structure in the data is accurately represented by the specified spatial model, the linearization of the latter will provide accurate estimates as long as the value of spatial dependence as measured by ρ is small, with more accurate estimates corresponding to smaller values of ρ. In addition, we expect the SMNL model to produce better estimates (with smaller bias and RMSE) than MNL in the presence of spatial dependence.

Two sets of simulation results for the equal probability choice scenario based on two different sample sizes are presented in Table 4-1 and Table 4-2. Table 4-1 reports the simulation results based on 1,000 replications for a sample of 1,000 observations. Table 4-2 reports results based on 500 replications for a sample of 5,000 observations. Both sets of results indicate that the performance of the MNL model in the estimation of the slope coefficients – as measured by the average bias and RMSE of the corresponding estimates – progressively deteriorates as the spatial dependence induced in the data increases. The SMNL estimator has a relatively similar performance to the MNL estimator regarding the estimation of the slopes, both in terms of average bias and RMSE, for the sample size of 1,000 observations (Table 4-1). Our expectation of the performance of SMNL estimator relative to MNL estimator regarding the estimation of the slopes as spatial dependence increases is only weakly confirmed by the results from the larger sample (Table 4-2). In this case, the SMNL estimator produces estimates that have smaller bias and RMSE compared to the corresponding estimates from the MNL model, however the difference in the average bias and RMSE of the slope coefficients produced by the two models is negligible.

An obvious disadvantage of the MNL model is that it does not produce an estimate of the spatial lag parameter ρ. This parameter is in itself of interest in practice as it indicates the presence of spatial interactions or spatial spillover effects. Regarding this parameter, the results exhibit the anticipated trend that as the spatial dependence induced in the data increases (i.e. for higher values of true ρ) so does the bias and RMSE of the estimated ρ. This trend is a byproduct of the model linearization. More specifically, our results indicate that ρ is accurately estimated by SMNL model when the

value of true ρ is zero — which is important since it correctly identifies the absence of spatial dependence — but also for values of ρ that are closer to zero (the linearization point). In particular, the SMNL estimator performs remarkably well in capturing the degree of spatial dependence induced in the simulated data for values of true ρ between 0.0 and 0.5. Another way to evaluate the performance of the SMNL estimator regarding the estimation of the spatial parameter is to compute the relative percentage bias, which expresses the average bias as a percentage of the true value of the spatial parameterρ[8]. In both sets of simulations the relative percentage bias of the estimated ρ when the value of true ρ is between 0.1 and 0.5 is relatively low and ranges between 0.3% (for $\rho = 0.1$) and 7.5% (for $\rho = 0.5$). The relative percentage bias becomes much higher for true values of ρ larger than 0.5, reaching up to 39% when ρ is set to 0.9.

The simulation results for the different choice probabilities scenario based on the sample sizes of 1,000 observations and 5,000 observations are presented in Table 4-3 and Table 4-4, respectively. These results exhibit similar general trends in terms of the deterioration of estimated slopes for both models as the induced spatial dependence (ρ) increases. However, The SMNL estimator outperforms the MNL estimator regarding the estimation of the slopes, both in terms of average bias and RMSE. In some instances, the average bias and RMSE of the estimates obtained by the SMNL model amount to less than half of the magnitude of the respective bias and RMSE measures for the MNL model. Thus, the estimation of the slopes deteriorates much faster when the model is estimated using MNL. These results confirm our expectations of the performance of SMNL estimator in the presence of spatial effects. We note, however, that in some

[8] $\text{Relative percentage bias} = \frac{\hat{\rho}-\rho}{\rho} x100\%.$

cases the slope coefficients estimated using the SMNL model still have considerable bias.

In terms of the estimation of the spatial parameter (ρ), these results exhibit similar trends to the previous scenario. As the induced spatial dependence (ρ) increases so does the bias and RMSE of the estimated ρ. As before, the SMNL model produces quite accurate estimates of ρ when the true value of ρ is set equal to zero and for small values of true ρ. In particular, the SMNL estimator performs well in capturing the degree of spatial dependence for true values of ρ between 0.0 and 0.5. In this case, the relative percentage bias of the estimated ρ ranges between 0.2% and 17% when the true value of the spatial parameter ρ is set between 0.1 and 0.5.The relative percentage bias of the estimated ρ increases up to 26% when the true value of ρ is set to 0.9. Higher accuracy (smaller bias) and precision (smaller RMSE) of the estimates can be obtained by increasing the sample size.

In general, the results from this Monte Carlo analysis indicate that the SMNL estimator captures well the degree of spatial dependence in the simulated data as long as the value of the induced spatial dependence as measured by ρ is small, hence closer to the linearization point. We further note that the SMNL model generally outperforms the MNL model with regard to the estimation of the slope coefficients, although the latter are still sometimes estimated with considerable bias. Judging from these simulation results, the linearization approach appears successful.

Determinants of Land-Use Choice

In this section, we follow the spatially-explicit approach described in the previous sections to analyze the factors that drive the land-use conversion process. We are

particularly interested in those factors that affect the pattern of new urban development as it occurs in rural-urban fringe areas. The typical pattern of new development in these areas throughout the U.S. is low-density and land-intensive, resulting in sprawling urban patterns. The area considered for this study is a rural-urban fringe county within the Cleveland, Ohio, metropolitan area. Medina County, located just south of the city of Cleveland and its suburbs, is typical of such development (Figure 4-1).

The parcel database for Medina County is comprised of data from the Medina County Auditor's office records. It includes information on land uses, parcel characteristics, major roads, and socio-economic indicators. The disaggregation scale of the data at the parcel level is appropriate for modeling the economic decisions of the individual landowners. In addition, the parcel-level data enables addressing the spatial heterogeneity of the landscape (e.g. soil type and quality) as well as local land-use policies (e.g. zoning).

To generate the parcel-level data, we use a Geographic Information System (GIS) that allows generating the relevant set of variables using the geocoded parcels and additional GIS data layers. The data layers include 1990 land use (Medina County and Cleveland State University), major roads (Ohio Department of Transportation), soil type (STATSGO), Census block group boundaries, and data from the U.S. Census of Population (U.S. Census Bureau). In addition, we use buffers to obtain neighborhood attributes that are likely to affect land conversion such as proportion of land in surrounding land uses, population and housing densities, etc.

Figure 4-2 shows changes in land-use patterns in Medina County from 1970 to 2000. The prevalent land development in this region is residential, accounting for more

than 85% of land development in the last 30 years. The pattern of urban development in Medina County is comprised of clusters of land tracts in similar land use, providing some preliminary visual evidence of positive spatial dependence. This type of evidence, nonetheless, is helpful for conceptualizing the structure of spatial dependence that needs to be specified and included in the model. Moreover, Figure 4-2 suggests that urban development in Medina County may involve divergent development processes for different urban land uses. For instance, residential development seems to have become more fragmented and dispersed, whereas the commercial and industrial land development has become more clustered over time. For this reason, we employ a model that encompasses multiple urban land uses.

The major urban center of our study area is Cleveland, so we measure proximity from each parcel centroid to the center of Cleveland via the major roads network (Totdiscle). Local markets are important for urban land uses; therefore, distance to nearest city (Disttonear) is included. To capture the potential disamenity effects of population on urban development, we measure the density of population in 1990 within the local neighborhood of each parcel (Popdens). The localized housing demand from the in-migration of urban residents in the region is captured by the proportion of houses in 1990 (Housedens).

To investigate whether surrounding land uses confer either positive or negative spillovers, we include three variables that measure the proportion of the surrounding land in residential (Reside), commercial (Commarea), and agricultural (Agarea) land uses, respectively. We also capture constraints to the density of development from large lot zoning with a dummy variable that is assigned a value of one if the minimum lot size

is zoned as three acres or greater and zero otherwise (Largelot). These variables measure, to an extent, some of the potential spatial dependence. Whether they suffice for that purpose can be seen from the statistical significance of the estimate of the spatial lag parameter in the SMNL model.

To control for differences in township-specific characteristics such as localized land-use policies, we construct dummy variables for each township and normalize estimation to the Homer and Sharon Townships, which are both very rural. Table 4-5 through Table 4-7 provide a description of the variables used in the model, descriptive statistics for those variables, and a summary of the proportion of the parcels in each land use category. The data set contains 9,760 parcels of which 5,991 parcels (61%) are in agricultural use; 2,917 parcels (30%) are in residential use; 572 parcels (6%) are in commercial use; and, 280 parcels (3%) are in industrial use. The average parcel is about 18 acres in size and is located about 100 miles from Cleveland.

Because the true specification of W is unknown, for our spatial analysis we construct four different weights matrices that impose varying assumptions about the extent and gradient of spatial dependence. All of them set the non-zero elements of the matrix to: $w_{ij} = 1/(dist)_{ij}^{f}$, which is the inverse Euclidean distance between locations i and j, and f a "friction" parameter. The first two specifications of W set the friction parameter to one and consider maximum cut-off distances—beyond which $w_{ij} = 0$ — at 800 and 1600 meters. The next two specifications set the friction parameter to two (inverse of the squared Euclidean distance) and employ the same cut-off distances. All matrices set $w_{ii} = 0$ and are row-standardized (rows sum to one).

The estimated parameters for the MNL model – our benchmark model – and the various specifications of the SMNL model are presented in Table 4-8, in which the base category chosen is the industrial land use. In general, the estimates from these models align with expectations. In addition, the SMNL estimates are reasonably robust across the different specifications. However, the sign of few of the coefficients changes occasionally across some of the SMNL model specifications. This point needs further investigation. A possible explanation for this discrepancy could be that some of the spatial weights matrices define neighborhoods that are rather large and include parcels in mixed land uses.

The estimates show that as distance to Cleveland (Totdiscle) increases the relative probabilities of parcels in agricultural, residential, and commercial uses decrease. So distance to the major urban center is an important factor. Local markets are also important to these land uses as measured by the distance to the nearest town (Disttonear). The estimates indicate that the probability of residential and agricultural land uses is higher relative to the industrial land use as distance to nearest town decreases. In contrast, the relative probability of the commercial land use increases with this distance. Another measure of proximity to local markets is per capita income (Percpinc). We would expect that the relative probability of agricultural land use decrease with a higher per capita income since agricultural land development tends to occur in economically depressed areas. However, our estimates do not suggest a significant per capita income effect.

Land characteristics of the parcel and characteristics of the surrounding area strongly affect the probability of agricultural use. For example, parcel size (Acres) has

the expected sign indicating that if parcel size increases, the probability of agricultural land uses is higher relative to the probability of industrial land use. The same holds for the commercial use. On the other hand, this relative probability is lower for the residential use as the parcel size increases. In addition, these estimates suggest that the relative probability between agricultural and industrial land use is higher when the majority of the area surrounding the parcel is in agricultural use (Agarea). Conversely, when urban uses surround the parcel (Reside and Commarea), the relative probability of agricultural use is lower.

The population density (Popdens) has a positive effect on the probability of commercial use relative to industrial use. Housing density (Housedens) is found to increase the relative probability of agricultural and commercial land use (relative to industrial use), but the magnitude of this effect is fairly small. In addition, the housing density negatively affects the relative probability of residential use, indicating a preference for lower density areas for this type of urban development. Interestingly, the effect of population and housing density on residential and commercial land uses becomes insignificant in some of the spatial model specifications, perhaps an indication that these neighborhood effects are adequately captured by the spatial specification of the model.

A binary variable representing large minimum-lot zoning (Largelot) is introduced as a land use policy variable. Estimates suggest that the minimum-lot size policy – minimum lot size zoned as three or more acres – decreases the relative probability of residential land use relative to the other urban land uses, implying that if the parcel is

subject to a restriction of a minimum lot size of three or more acres, the residential land use is less likely to occur.

Our estimates also include indicator variables for the township the parcel is located in, although their estimated parameters are not presented in the table for readability. The township parameter estimates indicate that a parcel located in any one of these townships is more likely to experience urban development than a parcel located in the townships to which we normalized the results: Homer Township and Sharon Township. This is expected since Homer Township and Sharon Township, unlike most of the other townships, are very rural and have experienced almost no urban growth.

Finally, and most importantly, the estimates of the spatial lag parameter (ρ) produced by the SMNL model vary from 0.26 to 0.49 (depending on the specification of W), and they are all highly statistically significant. This strongly suggests the presence of positive spatial interaction in land-use conversion decisions of landowners of neighboring parcels. More specifically, a high propensity of a given landowner to convert his land parcel to a particular land use positively affects the propensity of a landowner of an adjacent parcel to convert the latter to that specific land use, thus increasing the probability that the two neighboring parcels will convert to the same land use. This, in turn, implies clustering of land tracts that are in similar land use, a result that is in accordance with the spatial pattern of land uses for Medina County observed in Figure 4-2.

Since the discrete-choice models are nonlinear in parameters, the estimated coefficients for both the MNL and SMNL models do not have marginal effects

interpretation. Therefore, the marginal effects need to be computed. Note that in our spatial model the spatial effects are global, i.e. a change in the ith observation of the kth explanatory variable will have an impact on the expected probability of the event of interest (i.e the probability of conversion) not only for own region y_i (parcel i) but also the other regions y_j, $i \neq j$ (neighboring parcel j). This has implications when computing marginal effects as the marginal effects are no longer point estimates, but rather $(N \times N)$ matrices with the diagonal elements representing the direct effects of a change in the explanatory variable and the off-diagonal elements representing the indirect (or spillover) effects (LeSage and Pace, 2009). The marginal effects for our SMNL model are given by:

$$ME_{ik} = P_{ik}(\beta_k - \sum_l P_{il}\,\beta_l) \odot (I - \rho W)^{-1} \odot (\iota_i \frac{1}{\sigma_i}) \tag{4.16}$$

where ι_N denotes an $Nx1$ vector of ones and $\odot$ represents element-by-element (Hadamard) multiplication.

To provide scalar measures of the marginal effects, LeSage and Pace (2009) suggest to average over the diagonal elements to obtain a measure of the direct effects, average the row (or column) sums to produce a measure of the total effect, and take the difference between these two effects to get a measure of the indirect effects. More specifically, the marginal effects for the SMNL model are obtained as follows:

$$\begin{aligned} \overline{ME}\,(k)_{direct} &= N^{-1}\, tr(ME_k) \\ \overline{ME}\,(k)_{total} &= N^{-1}\, \iota_N'(ME_k)\iota_N \\ \overline{ME}(k)_{indirect} &= \overline{ME}\,(k)_{total} - \overline{ME}\,(k)_{direct} \end{aligned} \tag{4.17}$$

The estimates of the marginal effects for the MNL model and the different specifications of the SMNL model are presented in Table 4-9. The interpretation of a

marginal effect in a discrete-choice model is straightforward; it measures the change in the event probability associated with a change in the average observation for a given explanatory variable. The interpretation of the marginal effects produced by a spatial discrete-choice model is similar with the exception that now a change in a particular explanatory variable for a given observation (i) generates multiple effects which are summarized into a direct effect, indirect effect, and a total effect. The direct effect shows the impact of a one-unit change in a covariate on the event probability associated with the spatial unit of interest. The indirect effect measures the impact of a one-unit change in a covariate on the event probability associated with spatial units neighboring the spatial unit of interest. Naturally, the total effect, as the sum of these two effects, shows the average total impact in the event probability associated with a change in a particular observation (i.e. the spatial unit of interest). Note that a marginal effect produced from the MNL model for a specific choice alternative would be interpreted as a direct effect, which in this case also equals the total effect since the indirect effect in a non-spatial model is zero (LeSage and Pace, 2009).

Comparing the marginal effects produced by the MNL model with the corresponding direct effects from the SMNL model, the SMNL model produces marginal effects for all four land-use categories that are considerably larger in magnitude for about half of the explanatory variables. Note that the total effect of a marginal change in any of these explanatory variables in the probability of conversion of a parcel of land to each of these land-use categories is even larger if we consider the indirect effects. For the remaining half of the explanatory variables, the SMNL model produces larger direct effects (and total effects) for the commercial and industrial land-use categories. Thus,

the probabilistic impacts of the determinants of land-use changes obtained under the MNL model, especially for the commercial and industrial land-use categories, are likely understated. A major disadvantage of the MNL model is that it does not provide estimates of indirect impacts, which as the estimates from the SMNL model specifications show are in general statistically different from zero. Interestingly, the SMNL model specifications corresponding to the spatial weights matrices that define the neighborhood using a 1600 meter buffer produce estimates of indirect effects that are of nearly the same magnitude as the direct effects. In this case, changes occurring in a parcel affect as much the probability of conversion of surrounding parcels as they affect the probability of conversion of the own parcel. This, in turn, makes land conversion of neighboring land parcels to similar land uses much more likely.

The interpretation of the estimates of the marginal effects complies with those of the estimated coefficients. The size of the parcel has a negative effect on the probability of conversion to residential use, but a positive effect in the probability of conversion to agricultural, commercial, and industrial uses. This is expected since residential development generally occurs in smaller parcels. Distance to Cleveland negatively affects the probability of conversion of a parcel to residential and commercial use, and positively affects the probability of conversion to industrial use. Proximity to markets seems to matter more to the agricultural and the commercial uses.

The marginal effects associated with the proportion of land parcels in surrounding land uses correctly show that land parcels are likely to convert (or remain) to the land use that dominates the neighborhood. However, we would expect these effects to become insignificant in the spatial model specifications since the neighborhood is

modeled explicitly. Population density and housing density have a fairly small effect on the probability of conversion to all land uses. It seems like the spatial model is adequately capturing these neighborhood effects. The marginal effects for our policy variable suggest that if parcels are zoned to three or more acres they are more likely to stay in agricultural use. Lastly, across all specifications of the SMNL model, the direct effects and indirect effects have the same sign, a result that is consistent with the presence of positive spatial dependence.

On a final note, the SMNL model produces a NxN matrix of marginal effects for each of the determinants of land-use conversion to a given land use alternative, providing abundant information regarding the direct impacts of each of the determinants of land-use conversion on the probability of conversion of a particular land parcel as well as the indirect impacts that spill over from (or to) surrounding land parcels. The scalar measures reported in the section summarize these effects as averages over the whole sample. But, this is only one way to analyze these effects. As LeSage and Pace (2009) point out, if interest lies in a particular area (neighborhood) within the region of study rather than in the entire region, the NxN matrices of marginal effects can be appropriately partitioned to represent different neighborhoods and the marginal effects can be summarized in a similar manner. This is particularly helpful to identify hot spots of urban development.

Final Remarks

In this chapter, we estimate a spatially-explicit model of land-use conversion by employing a spatial multinomial logit model that explicitly accounts for spatial dependence in the land-use conversion decisions in the form of a spatial lag. We also assess the performance of the spatial multinomial logit (SMNL) estimator used to

estimate the model empirically through a Monte Carlo study. Simulation results indicate that the linearized spatial model provides a good approximation to the original spatial model and produces fairly accurate estimates of the degree of spatial dependence in the simulated data, provided the original spatial model is correctly specified and the spatial dependence is fairly low (i.e. the value of the spatial parameter is closer to the linearization point). Judging from these results, the linearization approach appears successful.

There are several advantages to the estimation methodology used in this chapter compared to other spatial methods and standard modeling methods. First, unlike the majority of spatial methodologies in the literature that are designed to model binary choice decisions, this methodology is designed to analyze spatially dependent multinomial choice decisions. Second, it is easy to implement even in large samples allowing us to estimate the model with disaggregated data at the parcel-level, a scale that is appropriate for modeling individual landowners' decisions. Third, it enables the estimation of the direct and indirect (spillover) effects of the determinants of land-use conversion, providing abundant information about the impacts of these determinants not only on the probability of conversion of a particular land parcel, but also on the probability of conversion of other land parcels located within a given proximity to the parcel of interest. In addition, it produces an estimate of the spatial lag parameter, which measures the strength of the interaction in land-use change decisions of landowners of nearby parcels. Knowledge of these effects, as will become more evident shortly, is of value to inform policy. Finally, as indicated by the results from the Monte Carlo study, the SMNL estimator performs well in capturing the spatial dependence in the data. A

drawback of this methodology is that since it is based on a linear approximation of the original spatial model, it does not perform equally well for empirical settings characterized by especially high spatial dependence.

The empirical results suggest that the location of new urban development is guided by a preference for lower density areas but in proximity to current urban development. This development trend is sometimes deterred by zoning policies. The most important insight from this application comes from the estimates of the spatial lag parameter, indicating that spatial dependence is an important feature of the land-use conversion process that needs to be taken into account when analyzing land-use conversion decisions for both empirical and policy reasons. Empirically, a land-use change model that ignores spatial dependence is likely to suffer from model misspecification. From a policy perspective, this result suggests that parcels do not develop in isolation, thus the land conversion that is likely to occur in a specific area influences land conversion in adjacent areas. For this reason, uncoordinated local land-use policies that are designed at a small scale such as subdivision regulation and zoning while attempting to locally manage growth may fragment urban development and result in suboptimal land-use patterns regionally. This is of major concern since it can be argued that the very policies that are designed to discourage sprawling urban development could be, at worst, causing it, and at best, enabling it (Attkisson, 2009; Batchis, 2010).

The power to design and enact land-use regulation in Ohio, like many other states that have not adopted a growth management plan, is delegated to the local governments (i.e. counties, townships, and municipalities) under the assumption that

they have a better understanding of problems of local concern. The local governments typically make land-use regulations following the statutory basis provided by the Standard State Zoning Enabling Act (SZEA) of 1926 and enact land regulations “for the purpose of promoting public health, safety, morals, or the general welfare of the community …in accordance with a comprehensive plan…”, although planning is not mandatory since the act failed to define “comprehensive plan” and the Standard Planning Act (SPA) that followed in 1928 was never integrated with the SZEA (Attkisson, 2009). The statutory provisions in the Ohio Revised Code that deal with the zoning powers of counties and townships are formulated in a virtually identical language (Hunt, 2001). In addition to zoning powers, the Ohio Revised Code grants similar powers to local governments to enact subdivision regulation. With such legislation in place, individual communities are justified to enact land-use regulations based entirely in their own self-interest such as to solve problems of a very local nature or advance their narrowly defined development goals.

Because of the spatial interactions that characterize the land-use conversion decisions as well as the spatial spillover effects associated with the land-use determinants including land-use regulation, the results from this study suggest that individual communities that act in self-interest fail to internalize the effects of their policies in the neighboring communities, resulting in suboptimal urbanization patterns at a regional level. For instance, a community that uses land-use controls (e.g. exclusionary zoning) to manage its urban areas will push inevitable urban growth to adjacent communities that do not have such regulation in place causing the region to sprawl. For this reason, coordination of local land-use policies at a regional level may

provide a way to internalize the externalities associated with urban growth and manage it more efficiently. In this regard, knowledge of the presence and the extent of spatial spillover effects can help inform the design of land use policies at an appropriate scale so as to achieve some vertical consistency between local and regional land-use policies and development goals.

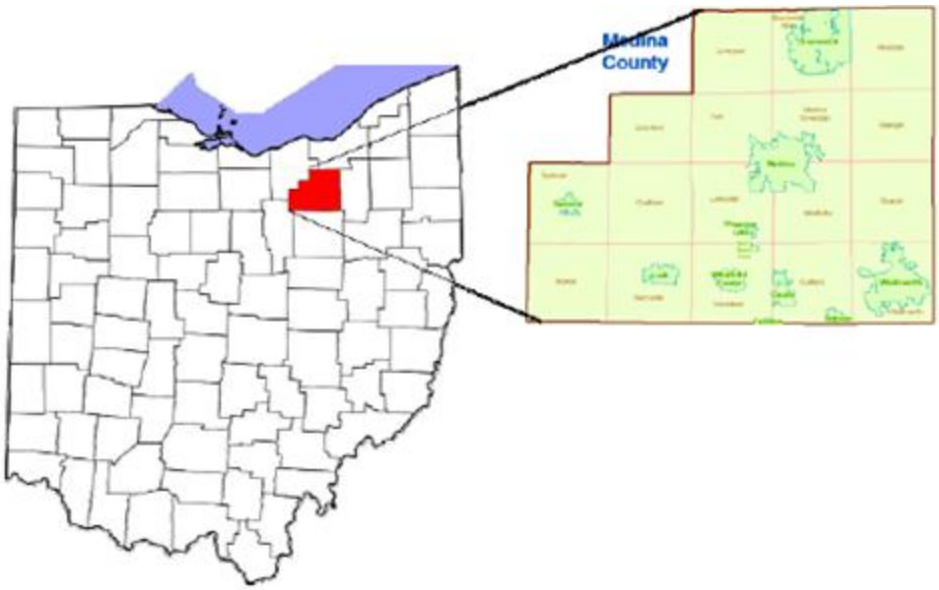


Figure 4-1. Medina County, Ohio

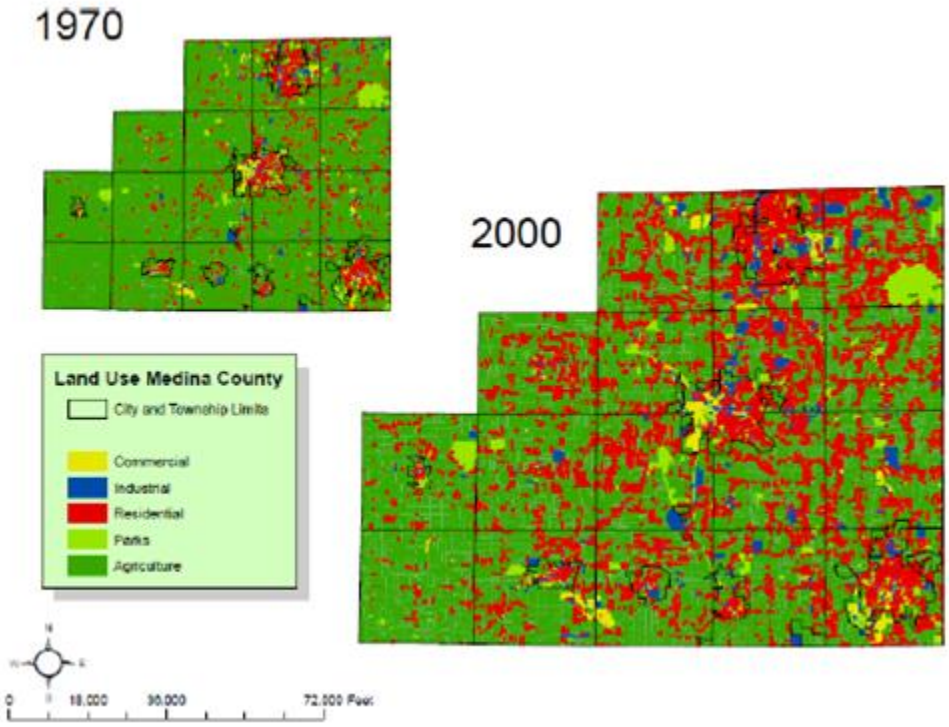


Figure 4-2. Land-use changes in Medina County, Ohio (1970-2000)

Table 4-1. Simulation results for the SMNL model with equal choice probabilities (Sample size: 1,000 observations)

	Standard Multinomial Logit			Spatial Multinomial Logit			
ρ	β1	β2	β3	β1	β2	β3	ρ
0							
Bias	0.0009	0.0006	0.0052	0.0010	0.0006	0.0052	-0.0020
RMSE	0.1591	0.1573	0.1634	0.1591	0.1573	0.1634	0.1949
0.1							
Bias	-0.0027	-0.0021	0.0020	-0.0029	-0.0024	0.0017	-0.0048
RMSE	0.1614	0.1576	0.1640	0.1614	0.1576	0.1640	0.1963
0.2							
Bias	-0.0113	-0.0101	-0.0071	-0.0119	-0.0107	-0.0077	-0.0040
RMSE	0.1616	0.1573	0.1646	0.1615	0.1573	0.1645	0.2002
0.3							
Bias	-0.0265	-0.0253	-0.0222	-0.0274	-0.0262	-0.0231	-0.0028
RMSE	0.1623	0.1585	0.1657	0.1623	0.1585	0.1656	0.2072
0.4							
Bias	-0.0486	-0.0482	-0.0444	-0.0497	-0.0494	-0.0456	0.0055
RMSE	0.1651	0.1615	0.1707	0.1652	0.1617	0.1708	0.2183
0.5							
Bias	-0.0769	-0.0774	-0.0730	-0.0783	-0.0788	-0.0745	0.0206
RMSE	0.1737	0.1722	0.1790	0.1741	0.1727	0.1794	0.2286
0.6							
Bias	-0.1180	-0.1171	-0.1136	-0.1196	-0.1187	-0.1153	0.0502
RMSE	0.1942	0.1927	0.1993	0.1950	0.1935	0.2000	0.2514
0.7							
Bias	-0.1712	-0.1709	-0.1676	-0.1730	-0.1726	-0.1694	0.0974
RMSE	0.2311	0.2290	0.2344	0.2323	0.2302	0.2355	0.2926
0.8							
Bias	-0.2500	-0.2471	-0.2444	-0.2518	-0.2489	-0.2463	0.1781
RMSE	0.2943	0.2899	0.2936	0.2957	0.2914	0.2950	0.3651
0.9							
Bias	-0.3749	-0.3718	-0.3688	-0.3766	-0.3736	-0.3706	0.3383
RMSE	0.4055	0.4018	0.4024	0.4070	0.4033	0.4039	0.5424

Note: Results are based on a sample with 1,000 observations and 1,000 replications. The slope's true value is 1 in all cases.

Table 4-2. Simulation results for the SMNL model with equal choice probabilities (Sample size: 5,000 observations)

	Standard Multinomial Logit			Spatial Multinomial Logit			
ρ	β1	β2	β3	β1	β2	β3	ρ
0							
Bias	-0.0018	0.0005	0.0016	-0.0018	0.0005	0.0016	0.0008
RMSE	0.0713	0.0731	0.0725	0.0713	0.0731	0.0725	0.0794
0.1							
Bias	-0.0058	-0.0026	-0.0013	-0.0056	-0.0024	-0.0011	0.0005
RMSE	0.0709	0.0736	0.0725	0.0709	0.0736	0.0726	0.0798
0.2							
Bias	-0.0136	-0.0101	-0.0097	-0.0132	-0.0097	-0.0093	0.0027
RMSE	0.0737	0.0748	0.0733	0.0736	0.0747	0.0732	0.0831
0.3							
Bias	-0.0276	-0.0241	-0.0243	-0.0271	-0.0235	-0.0238	0.0075
RMSE	0.0778	0.0777	0.0762	0.0777	0.0775	0.0761	0.0867
0.4							
Bias	-0.0495	-0.0455	-0.0461	-0.0488	-0.0448	-0.0454	0.0195
RMSE	0.0881	0.0866	0.0858	0.0877	0.0862	0.0854	0.0912
0.5							
Bias	-0.0774	-0.0743	-0.0750	-0.0765	-0.0733	-0.0741	0.0373
RMSE	0.1056	0.1053	0.1039	0.1050	0.1047	0.1032	0.1012
0.6							
Bias	-0.1165	-0.1131	-0.1141	-0.1154	-0.1119	-0.1130	0.0678
RMSE	0.1368	0.1352	0.1348	0.1359	0.1343	0.1339	0.1220
0.7							
Bias	-0.1687	-0.1662	-0.1672	-0.1674	-0.1649	-0.1659	0.1160
RMSE	0.1830	0.1821	0.1818	0.1819	0.1810	0.1806	0.1617
0.8							
Bias	-0.2462	-0.2412	-0.2434	-0.2448	-0.2398	-0.2420	0.1980
RMSE	0.2562	0.2521	0.2534	0.2550	0.2508	0.2520	0.2379
0.9							
Bias	-0.3676	-0.3635	-0.3656	-0.3662	-0.3620	-0.3642	0.3491
RMSE	0.3742	0.3706	0.3723	0.3728	0.3692	0.3710	0.3899

Note: Results are based on a sample with 5,000 observations and 500 replications. The slope's true value is 1 in all cases.

Table 4-3. Simulation results for the SMNL model with different choice probabilities (Sample size: 1,000 observations)

		Standard Multinomial Logit			Spatial Multinomial Logit			
ρ		β1	β2	β3	β1	β2	β3	ρ
0.0	True Coeff	1.5980	0.9098	0.4026				
	Bias	0.0033	0.0015	-0.0006	-0.0070	-0.0009	-0.0011	0.0109
	RMSE	0.1148	0.1264	0.1400	0.1828	0.1324	0.1401	0.1792
0.1	True Coeff	1.4490	0.8250	0.3650				
	Bias	0.1467	0.0873	0.0386	0.0543	0.0611	0.0281	0.0118
	RMSE	0.1861	0.1528	0.1450	0.1933	0.1437	0.1409	0.1805
0.2	True Coeff	1.3181	0.7504	0.3321				
	Bias	0.2716	0.1614	0.0728	0.1084	0.1145	0.0535	-0.0005
	RMSE	0.2948	0.2035	0.1582	0.2154	0.1713	0.1470	0.1807
0.3	True Coeff	1.2005	0.6835	0.3025				
	Bias	0.3832	0.2273	0.1035	0.1566	0.1618	0.0761	-0.0212
	RMSE	0.3999	0.2592	0.1741	0.2452	0.2058	0.1552	0.1850
0.4	True Coeff	1.0928	0.6221	0.2753				
	Bias	0.4855	0.2879	0.1312	0.1980	0.2043	0.0962	-0.0446
	RMSE	0.4986	0.3136	0.1912	0.2739	0.2399	0.1642	0.1907
0.5	True Coeff	0.9915	0.5645	0.2498				
	Bias	0.5798	0.3436	0.1566	0.2366	0.2434	0.1144	-0.0732
	RMSE	0.5908	0.3655	0.2095	0.3034	0.2735	0.1746	0.2005
0.6	True Coeff	0.8935	0.5087	0.2251				
	Bias	0.6704	0.3964	0.1805	0.2758	0.2809	0.1317	-0.1064
	RMSE	0.6801	0.4158	0.2285	0.3354	0.3074	0.1863	0.2152
0.7	True Coeff	0.7946	0.4524	0.2002				
	Bias	0.7605	0.4489	0.2041	0.3189	0.3192	0.1490	-0.1440
	RMSE	0.7690	0.4661	0.2473	0.3721	0.3425	0.1979	0.2384
0.8	True Coeff	0.6881	0.3917	0.1733				
	Bias	0.8574	0.5053	0.2289	0.3700	0.3619	0.1677	-0.1825
	RMSE	0.8649	0.5207	0.2681	0.4171	0.3822	0.2117	0.2656
0.9	True Coeff	0.5559	0.3165	0.1400				
	Bias	0.9781	0.5751	0.2600	0.4451	0.4180	0.1926	-0.2191
	RMSE	0.9846	0.5886	0.2946	0.4856	0.4356	0.2309	0.2946

Note: Results are based on a sample with 1,000 observations and 1,000 replications. The slopes' true value $\beta_k(\rho)$ changes with ρ. The choice probabilities are set as follows: $P_0 = 0.10$, $P_1 = 0.50$, $P_2 = 0.25$, and $P_3 = 0.15$.

Table 4-4. Simulation results for the SMNL model with different choice probabilities (Sample size: 5,000 observations)

		Standard Multinomial Logit			Spatial Multinomial Logit			
ρ		β1	β2	β3	β1	β2	β3	ρ
0.0	True Coeff	1.6097	0.9164	0.4055				
	Bias	0.0026	0.0021	0.0032	0.0014	0.0018	0.0030	0.0012
	RMSE	0.0500	0.0539	0.0587	0.0791	0.0570	0.0589	0.0742
0.1	True Coeff	1.4597	0.8310	0.3677				
	Bias	0.1469	0.0883	0.0426	0.0643	0.0643	0.0324	0.0000
	RMSE	0.1550	0.1034	0.0723	0.1028	0.0855	0.0663	0.0762
0.2	True Coeff	1.3278	0.7559	0.3345				
	Bias	0.2724	0.1632	0.0765	0.1177	0.1181	0.0572	-0.0114
	RMSE	0.2769	0.1719	0.0968	0.1432	0.1305	0.0814	0.0783
0.3	True Coeff	1.2094	0.6885	0.3047				
	Bias	0.3843	0.2294	0.1065	0.1645	0.1650	0.0788	-0.0307
	RMSE	0.3876	0.2357	0.1223	0.1839	0.1740	0.0980	0.0838
0.4	True Coeff	1.1008	0.6267	0.2773				
	Bias	0.4854	0.2885	0.1332	0.2057	0.2064	0.0979	-0.0557
	RMSE	0.4881	0.2936	0.1463	0.2216	0.2136	0.1137	0.0964
0.5	True Coeff	0.9987	0.5686	0.2516				
	Bias	0.5802	0.3445	0.1581	0.2457	0.2460	0.1157	-0.0862
	RMSE	0.5824	0.3488	0.1691	0.2592	0.2519	0.1290	0.1172
0.6	True Coeff	0.8999	0.5123	0.2267				
	Bias	0.6711	0.3976	0.1822	0.2849	0.2836	0.1331	-0.1197
	RMSE	0.6730	0.4013	0.1919	0.2970	0.2886	0.1447	0.1447
0.7	True Coeff	0.8002	0.4556	0.2016				
	Bias	0.7623	0.4509	0.2064	0.3273	0.3222	0.1508	-0.1558
	RMSE	0.7640	0.4541	0.2151	0.3381	0.3266	0.1611	0.1766
0.8	True Coeff	0.6928	0.3944	0.1745				
	Bias	0.8601	0.5078	0.2319	0.3788	0.3651	0.1700	-0.1939
	RMSE	0.8616	0.5106	0.2395	0.3882	0.3688	0.1789	0.2117
0.9	True Coeff	0.5595	0.3185	0.1410				
	Bias	0.9821	0.5784	0.2641	0.4548	0.4218	0.1958	-0.2311
	RMSE	0.9834	0.5808	0.2708	0.4628	0.4251	0.2035	0.2468

Note: Results are based on a sample with 5,000 observations and 500 replications. The slopes' true value $\beta_k(\rho)$ changes with ρ. The choice probabilities are set as follows: $P_0 = 0.10$, $P_1 = 0.50$, $P_2 = 0.25$, and $P_3 = 0.15$.

Table 4-5. Land-use conversion model: variable description

Variable	Description
Landuse	1990 land use (=0 if agricultural, 1 if residential, 2 if commercial, and 3 if industrial)
Acres	Parcel size in acres
Percpinc	Per capita income
Totdiscle	Distance of each parcel center in meters from the center of Cleveland
Disttonear	Distance of each parcel center in meters to the nearest city
Popdens	1990 population density within each parcel's neighborhood
Housedens	Proportion of houses in 1990
Reside	Proportion of land in residential use
Commarea	Proportion of land in commercial use
Agarea	Proportion of land in agricultural use
Largelot	Largelot (=1 if min lot size zoned as >= 3 acres, 0 otherwise)

Table 4-6. Land-use conversion model: variable descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Landuse	0.50	0.73	0.00	3.00
Acres	17.60	28.55	<1.00	1,150.56
Percpinc	14,130.06	2,298.59	5,831.00	19,952.00
Totdistcle	166,294.10	38,839.08	222.14	261,608.00
Disttonear	17,663.67	8,494.32	192.25	40,377.11
Popdens	256.77	379.57	4.81	3,798.63
Housden	86.91	126.47	1.95	1,315.74
Reside	0.28	0.17	0.00	0.99
Comarea	0.04	0.07	0.00	0.51
Agarea	0.64	0.21	0.00	1.00
Largelot	0.16	0.37	0.00	1.00

Table 4-7. Proportion of parcels in each land-use category

LandUse	Parcels	Share (%)
Agricultural	5,991	61.38
Residential	2,917	29.89
Commercial	572	5.86
Industrial	280	2.87
Total	9,760	100.00

Table 4-8. Estimated coefficients of the land-use change model

			SMNL Models							
	MNL Model		W_800_f=1		W_800_f=2		W_1600_f=1		W_1600_f=2	
Variables	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.
acres:										
β*ag*	0.080	0.009	0.052	0.007	0.048	0.007	0.025	0.005	0.032	0.005
βres	-0.253	0.014	-0.819	0.095	-0.762	0.101	-0.432	0.107	-0.988	0.089
β*com*	0.068	0.009	0.061	0.009	0.068	0.008	0.083	0.007	0.044	0.009
totdiscle:										
(x10,000) β*ag*	-0.230	0.061	-0.242	0.374	-0.408	0.391	-0.095	0.287	-0.013	0.335
βres	-0.235	0.067	-1.328	0.623	-1.988	0.660	-4.488	0.593	-3.122	0.578
β*com*	-0.319	0.067	-2.499	0.682	-1.444	0.679	-0.837	0.73	-4.658	1.043
disttonear:										
(x10,000) *βag*	0.425	0.163	6.232	1.138	8.044	1.322	0.218	0.987	-4.865	0.907
βres	0.278	0.170	-1.821	1.221	-0.607	1.320	-1.903	1.381	3.323	1.046
β*com*	1.060	0.192	9.315	4.177	3.163	4.428	2.848	2.902	11.701	2.709
agarea:										
β*ag*	1.797	0.807	5.277	0.966	5.702	1.020	-1.757	0.902	-1.651	0.999
βres	1.848	0.834	1.695	1.340	1.604	1.358	3.455	1.061	6.145	1.216
β*com*	2.074	1.008	11.898	4.779	6.018	5.334	5.811	3.809	1.260	3.180
reside:										
β*ag*	-1.906	0.929	5.436	2.233	12.814	2.245	6.886	1.694	13.678	1.561
βres	4.234	0.954	7.542	3.016	0.633	2.912	7.856	1.955	14.525	2.288
β*com*	0.283	1.141	-12.219	4.006	-10.628	4.532	-5.295	3.669	-27.136	4.419
commarea:										
β*ag*	-0.016	1.495	10.169	2.742	13.140	3.038	-0.027	1.982	-2.075	2.199
βres	3.753	1.509	-3.953	3.710	-7.230	3.678	-5.812	3.338	6.997	3.598
β*com*	4.470	1.683	-0.357	9.851	-10.467	11.174	-13.157	5.021	-14.925	4.417

Table 4-8. Continued

			SMNL Models							
	MNL Model		W_800_f=1		W_1800_f=2		W_1600_f=1		W_1600_f=2	
Variables	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.	Est.	St. Err.
popdens:										
β*ag*	0.006	0.001	-0.004	0.004	-0.018	0.005	-0.003	0.003	0.018	0.004
βres	0.003	0.001	0.000	0.004	0.004	0.004	0.017	0.003	0.004	0.004
β*com*	0.011	0.002	0.023	0.011	0.016	0.013	0.039	0.012	0.041	0.010
housedens:										
β*ag*	-0.018	0.003	0.016	0.012	0.048	0.015	0.005	0.008	-0.050	0.011
βres	-0.011	0.003	-0.009	0.011	-0.013	0.012	-0.061	0.010	-0.021	0.011
β*com*	-0.034	0.005	-0.017	0.030	-0.008	0.037	-0.052	0.035	-0.088	0.028
percpinc:										
(x10,000) β*ag*	-0.383	0.418	0.936	1.765	2.198	1.653	4.909	1.445	8.785	1.612
βres	-0.011	0.429	2.051	1.865	-0.519	1.780	-0.139	1.431	0.469	1.561
β*com*	-1.271	0.494	-6.162	4.366	0.963	4.291	-5.091	4.976	-46.482	8.486
largelot:										
β*ag*	0.758	0.233	0.661	0.352	1.571	0.392	-1.459	0.279	-2.511	0.310
βres	0.284	0.245	0.544	0.299	0.549	0.302	-1.018	0.360	0.404	0.332
β*com*	0.881	0.261	3.990	1.031	2.279	0.932	1.215	0.783	-1.391	0.858
rho	--	--	0.261	0.027	0.331	0.027	0.487	0.04	0.400	0.036

Note: Sample size is 9,760 parcels. All models include indicator variables for the township in which the parcel resides. The columns for the SMNL model correspond to different specifications of the spatial weights matrix W that vary the cut-off distance (800 and 1600) and the friction parameter (f=1 or 2). See text for details.

Table 4-9. Marginal effects of the estimated coefficients of the land-use change model

		SMNL Models											
MNL Model		W_800_f=1			W_800_f=2			W_1600_f=1			W_1600_f=2		
Marginal Effects		Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
acres:													
ag	0.0273	0.0079	0.0016	0.0095	0.0048	0.0015	0.0063	0.0033	0.0025	0.0058	0.0049	0.0027	0.0075
	(0.0007)	(0.0011)	(0.0003)	(0.0014)	(0.0010)	(0.0003)	(0.0012)	(0.0003)	(0.0002)	(0.0005)	(0.0003)	(0.0002)	(0.0005)
res	-0.0322	-0.0167	-0.0033	-0.0201	-0.0142	-0.0039	-0.0181	-0.0061	-0.0046	-0.0107	-0.0089	-0.0049	-0.0138
	(0.0009)	(0.0016)	(0.0004)	(0.0020)	(0.0016)	(0.0004)	(0.0021)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0002)	(0.0005)
com	0.0034	0.0051	0.0010	0.0062	0.0057	0.0015	0.0072	0.0022	0.0016	0.0038	0.0019	0.0010	0.0029
	(0.0002)	(0.0004)	(0.0001)	(0.0004)	(0.0006)	(0.0002)	(0.0007)	(0.0002)	(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0003)
ind	0.0015	0.0037	0.0007	0.0044	0.0036	0.0009	0.0046	0.0006	0.0005	0.0011	0.0022	0.0012	0.0034
	(0.0002)	(0.0009)	(0.0002)	(0.0011)	(0.0004)	(0.0001)	(0.0005)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)
totdiscle (x 10,000):													
ag	0.0005	0.0223	0.0054	0.0277	0.0157	0.0052	0.0209	0.0406	0.0303	0.0708	0.0284	0.0160	0.0444
	(0.0032)	(0.0195)	(0.0057)	(0.0245)	(0.0243)	(0.0077)	(0.0318)	(0.0411)	(0.0314)	(0.0724)	(0.0249)	(0.0155)	(0.0403)
res	-0.0020	-0.0103	-0.0022	-0.0125	-0.0239	-0.0068	-0.0307	-0.0567	-0.0421	-0.0988	-0.0197	-0.0110	-0.0307
	(0.0033)	(0.0021)	(0.0005)	(0.0025)	(0.0039)	(0.0011)	(0.0049)	(0.0022)	(0.0017)	(0.0039)	(0.0068)	(0.0029)	(0.0094)
com	-0.0048	-0.0350	-0.0090	-0.0439	-0.0147	-0.0056	-0.0203	-0.0005	-0.0006	-0.0011	-0.0308	-0.0179	-0.0487
	(0.0016)	(0.0042)	(0.0012)	(0.0053)	(0.0017)	(0.0007)	(0.0023)	(0.0053)	(0.0041)	(0.0094)	(0.0029)	(0.0015)	(0.0044)
ind	0.0063	0.0229	0.0057	0.0287	0.0229	0.0071	0.0300	0.0166	0.0125	0.0291	0.0221	0.0129	0.0350
	(0.0018)	(0.0028)	(0.0007)	(0.0034)	(0.0016)	(0.0005)	(0.0021)	(0.0023)	(0.0018)	(0.0041)	(0.0013)	(0.0008)	(0.0021)
disttonear (x 10,000):													
ag	-0.0070	0.1015	0.0220	0.1235	0.1612	0.0543	0.2155	-0.0067	-0.0054	-0.0121	-0.1385	-0.0793	-0.2177
	(0.0092)	(0.0126)	(0.0030)	(0.0156)	(0.0105)	(0.0037)	(0.0140)	(0.0055)	(0.0041)	(0.0095)	(0.0659)	(0.0416)	(0.1072)
res	-0.0174	-0.1395	-0.0274	-0.1669	-0.0766	-0.0221	-0.0987	-0.0344	-0.0255	-0.0600	0.0317	0.0176	0.0493
	(0.0083)	(0.1080)	(0.2990)	(1.3591)	(0.1245)	(0.0357)	(0.1601)	(0.0287)	(0.0215)	(0.0501)	(0.0244)	(0.0183)	(0.0418)
com	0.0349	0.1315	0.0308	0.1624	-0.0055	-0.0047	-0.0101	0.0476	0.0360	0.0836	0.1003	0.0579	0.1582
	(0.0061)	(0.0099)	(0.0032)	(0.0129)	(0.0949)	(0.0314)	(0.1256)	(0.0611)	(0.0461)	(0.1071)	(0.0071)	(0.0036)	(0.0106)
ind	-0.0106	-0.0935	-0.0254	-0.1189	-0.0791	-0.0275	-0.1066	-0.0065	-0.0051	-0.0115	0.0065	0.0038	0.0103
	(0.0041)	(0.0052)	(0.0013)	(0.0065)	(0.0386)	(0.0135)	(0.0519)	(0.0177)	(0.0133)	(0.0310)	(0.0039)	(0.0023)	(0.0062)

Table 4-9. Continued

		SMNL Models											
MNL Model		W_800_f=1			W_800_f=2			W_1600_f=1			W_1600_f=2		
Marginal Effects		Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
agarea:													
ag	0.0112	0.0313	0.0059	0.0372	0.0664	0.0219	0.0884	-0.1400	-0.1055	-0.2455	-0.0611	-0.0343	-0.0953
	(0.0494)	(0.0075)	(0.0017)	(0.0091)	(0.0062)	(0.0022)	(0.0084)	(0.1092)	(0.0822)	(0.1912)	(0.0343)	(0.0198)	(0.0539)
res	0.0216	-0.0767	-0.0147	-0.0915	-0.0417	-0.0115	-0.0532	0.0445	0.0331	0.0776	0.0595	0.0327	0.0922
	(0.0435)	(0.0664)	(0.0156)	(0.0782)	(0.0767)	(0.0214)	(0.0977)	(0.0037)	(0.0028)	(0.0066)	(0.0218)	(0.0149)	(0.0363)
com	0.0161	0.1641	0.0400	0.2042	0.0643	0.0199	0.0842	0.0999	0.0761	0.1760	0.0047	0.0031	0.0078
	(0.0337)	(0.0140)	(0.0043)	(0.0181)	(0.0100)	(0.0035)	(0.0134)	(0.1144)	(0.0842)	(0.1985)	(0.0307)	(0.0201)	(0.0507)
ind	-0.0489	-0.1187	-0.0312	-0.1499	-0.0891	-0.0302	-0.1193	-0.0044	-0.0037	-0.0081	-0.0031	-0.0016	-0.0047
	(0.0211)	(0.0939)	(0.0253)	(0.1185)	(0.0459)	(0.0180)	(0.0635)	(0.0029)	(0.0023)	(0.0052)	(0.0021)	(0.0012)	(0.0034)
reside:													
ag	-0.5831	0.1453	0.0405	0.1858	0.3974	0.1375	0.5349	0.2064	0.1562	0.3627	0.2638	0.1539	0.4178
	(0.0557)	(0.0097)	(0.0032)	(0.0126)	(0.0313)	(0.0108)	(0.0416)	(0.0133)	(0.0097)	(0.0229)	(0.0126)	(0.0085)	(0.0210)
res	0.5643	0.1661	0.0329	0.1991	0.0037	0.0030	0.0067	0.0643	0.0476	0.1118	0.1115	0.0611	0.1726
	(0.0479)	(0.0101)	(0.0035)	(0.0133)	(0.0145)	(0.0055)	(0.0200)	(0.0058)	(0.0042)	(0.0100)	(0.0076)	(0.0068)	(0.0142)
com	0.0383	-0.2910	-0.0704	-0.3614	-0.3541	-0.1191	-0.4732	-0.1725	-0.1307	-0.3033	-0.2771	-0.1584	-0.4354
	(0.0337)	(0.0218)	(0.0070)	(0.0285)	(0.0272)	(0.0100)	(0.0364)	(0.2112)	(0.1548)	(0.3658)	(0.0141)	(0.0061)	(0.0199)
ind	-0.0195	-0.0204	-0.0030	-0.0234	-0.0471	-0.0154	-0.0624	-0.0981	-0.0731	-0.1712	-0.0983	-0.0566	-0.1549
	(0.0238)	(0.0165)	(0.0047)	(0.0210)	(0.0166)	(0.0067)	(0.0233)	(0.0142)	(0.0109)	(0.0251)	(0.0817)	(0.0371)	(0.1177)
commarea:													
ag	-0.4619	0.2776	0.0657	0.3433	0.4486	0.1534	0.6021	0.1760	0.1332	0.3092	-0.0218	-0.0113	-0.0331
	(0.0796)	(0.0259)	(0.0071)	(0.0329)	(0.0342)	(0.0117)	(0.0453)	(0.2239)	(0.1666)	(0.3903)	(0.0788)	(0.0500)	(0.1287)
res	0.3224	-0.1627	-0.0323	-0.1950	-0.1385	-0.0423	-0.1808	-0.0426	-0.0318	-0.0744	0.0958	0.0523	0.1481
	(0.0667)	(0.0211)	(0.0049)	(0.0260)	(0.0305)	(0.0093)	(0.0398)	(0.0789)	(0.0588)	(0.1376)	(0.0569)	(0.0417)	(0.0970)
com	0.1865	-0.0526	-0.0158	-0.0684	-0.3049	-0.1069	-0.4118	-0.1915	-0.1457	-0.3372	-0.1207	-0.0688	-0.1895
	(0.0467)	(0.1672)	(0.0427)	(0.2091)	(0.3219)	(0.1144)	(0.4309)	(0.0205)	(0.0148)	(0.0353)	(0.0663)	(0.0358)	(0.0985)
ind	-0.0471	-0.0623	-0.0176	-0.0799	-0.0053	-0.0042	-0.0094	0.0580	0.0443	0.1024	0.0467	0.0278	0.0745
	(0.0389)	(0.0859)	(0.0254)	(0.1111)	(0.0204)	(0.0076)	(0.0280)	(0.0106)	(0.0081)	(0.0186)	(0.0030)	(0.0019)	(0.0048)

Table 4-9. Continued

		SMNL Models											
MNL Model		W_800_f=1			W_800_f=2			W_1600_f=1			W_1600_f=2		
Marginal Effects		Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
popdens:													
ag	0.0001	-0.0002	-0.0001	-0.0003	-0.0006	-0.0002	-0.0008	-0.0006	-0.0005	-0.0011	0.0002	0.0001	0.0003
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0004)	(0.0001)	(0.0005)	(0.0006)	(0.0005)	(0.0011)	(0.0002)	(0.0001)	(0.0003)
res	-0.0003	-0.0001	-0.0000	-0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0003	-0.0001	-0.0001	-0.0002
	(0.0001)	(0.0001)	(0.0000)	(0.0002)	(0.0003)	(0.0001)	(0.0004)	(0.0002)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0001)
com	0.0003	0.0004	0.0001	0.0005	0.0005	0.0002	0.0007	0.0006	0.0005	0.0010	0.0003	0.0001	0.0004
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0004)	(0.0002)	(0.0006)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
ind	-0.0001	-0.0001	-0.0000	-0.0001	0.0000	0.0000	0.0000	-0.0001	-0.0001	-0.0002	-0.0003	-0.0002	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0003)	(0.0001)	(0.0004)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
housedens:													
ag	-0.0002	0.0006	0.0001	0.0007	0.0013	0.0004	0.0017	0.0011	0.0009	0.0020	-0.0005	-0.0003	-0.0008
	(0.0003)	(0.0005)	(0.0001)	(0.0006)	(0.0008)	(0.0003)	(0.0011)	(0.0011)	(0.0008)	(0.0020)	(0.0004)	(0.0003)	(0.0007)
res	0.0007	-0.0002	-0.0000	-0.0003	-0.0005	-0.0001	-0.0006	-0.0007	-0.0005	-0.0012	0.0002	0.0001	0.0003
	(0.0002)	(0.0004)	(0.0001)	(0.0005)	(0.0009)	(0.0003)	(0.0012)	(0.0000)	(0.0000)	(0.0001)	(0.0002)	(0.0001)	(0.0003)
com	-0.0009	-0.0004	-0.0001	-0.0005	-0.0006	-0.0002	-0.0008	-0.0007	-0.0005	-0.0012	-0.0005	-0.0003	-0.0008
	(0.0002)	(0.0007)	(0.0002)	(0.0009)	(0.0008)	(0.0003)	(0.0011)	(0.0010)	(0.0007)	(0.0017)	(0.0003)	(0.0002)	(0.0005)
ind	0.0004	0.0000	0.0000	0.0000	-0.0003	-0.0001	-0.0004	0.0003	0.0002	0.0005	0.0008	0.0005	0.0013
	(0.0001)	(0.0004)	(0.0001)	(0.0005)	(0.0005)	(0.0002)	(0.0006)	(0.0004)	(0.0003)	(0.0008)	(0.0002)	(0.0002)	(0.0004)
percpinc (x 10,000):													
ag	0.0000	0.0456	0.0130	0.0586	0.0451	0.0151	0.0603	0.2053	0.1548	0.3600	0.3003	0.1733	0.4736
	(0.0235)	(0.0681)	(0.0174)	(0.0830)	(0.0574)	(0.0196)	(0.0768)	(0.1156)	(0.0829)	(0.1982)	(0.2091)	(0.1381)	(0.3468)
res	0.0431	0.0652	0.0128	0.0780	-0.0278	-0.0079	-0.0357	-0.0270	-0.0201	-0.0471	0.0418	0.0222	0.0640
	(0.0206)	(0.0560)	(0.0174)	(0.0723)	(0.0609)	(0.0168)	(0.0777)	(0.0540)	(0.0398)	(0.0938)	(0.0707)	(0.0682)	(0.1349)
com	-0.0505	-0.1253	-0.0304	-0.1557	0.0028	0.0000	0.0028	-0.1287	-0.0978	-0.2266	-0.3882	-0.2229	-0.6111
	(0.0149)	(0.1467)	(0.0422)	(0.1880)	(0.0085)	(0.0028)	(0.0113)	(0.1979)	(0.1451)	(0.3429)	(0.0237)	(0.0108)	(0.0343)
ind	0.0074	0.0145	0.0046	0.0191	-0.0201	-0.0072	-0.0273	-0.0496	-0.0369	-0.0864	0.0461	0.0274	0.0735
	(0.0109)	(0.0405)	(0.0118)	(0.0520)	(0.0259)	(0.0091)	(0.0350)	(0.6615)	(0.5077)	(1.1692)	(0.0172)	(0.0097)	(0.2688)

Table 4-9. Continued

		SMNL Models											
MNL Model		W_800_f=1			W_800_f=2			W_1600_f=1			W_1600_f=2		
Marginal Effects		Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
largelot:													
ag	0.0426	-0.0156	-0.0043	-0.0199	0.0118	0.0037	0.0155	-0.0497	-0.0376	-0.0873	-0.0423	-0.0242	-0.0665
	(0.0127)	(0.0191)	(0.0048)	(0.0235)	(0.0015)	(0.0005)	(0.0020)	(0.0258)	(0.0185)	(0.0443)	(0.0226)	(0.0171)	(0.0497)
res	-0.0402	-0.0166	-0.0032	-0.0198	-0.0135	-0.0036	-0.0171	-0.0053	-0.0039	-0.0091	0.0172	0.0094	0.0265
	(0.0122)	(0.0216)	(0.0047)	(0.0252)	(0.0212)	(0.0061)	(0.0271)	(0.0011)	(0.0008)	(0.0020)	(0.0756)	(0.0503)	(0.1246)
com	0.0129	0.0626	0.0155	0.0780	0.0304	0.0096	0.0400	0.0363	0.0276	0.0639	-0.0036	-0.0020	-0.0056
	(0.0067)	(0.0052)	(0.0016)	(0.0068)	(0.0031)	(0.0013)	(0.0043)	(0.0447)	(0.0326)	(0.0773)	(0.0094)	(0.0073)	(0.0166)
ind	-0.0152	-0.0303	-0.0080	-0.0383	-0.0287	-0.0097	-0.0384	0.0187	0.0139	0.0326	0.0287	0.0168	0.0455
	(0.0062)	(0.0353)	(0.0102)	(0.0453)	(0.0200)	(0.0079)	(0.0278)	(0.0029)	(0.0022)	(0.0051)	(0.0111)	(0.0070)	(0.0180)

Note: The standard errors of the estimated marginal effects presented in parentheses are obtained using the Delta method.

CHAPTER 5
NEIGHBORHOOD INFLUENCE AND PUBLIC POLICY: ADOPTION OF OPEN ENROLLMENT POLICIES BY SCHOOL DISTRICTS IN OHIO

Background

Public school reform has been at the heart of public debate in the United States (U.S.) since the early 1980s concerning issues of student academic performance and dropout rates, national academic standards, school governance and funding, equity and diversity, safety and desegregation (Crepage, 1999). Of many strategies considered to bring about school reform in the public school system, public school choice received considerable attention. Public school choice refers to policies that provide parents with an opportunity to select an educational program or a public school for their children to attend tuition-free other than the neighborhood school to which they would be normally assigned based on the district of residence[1]. Eight basic types of school choice policies have been proposed and implemented by various states to address public education challenges ranging from desegregation of schools to improvement of education practices: magnet schools, post-secondary enrollment programs, dropout prevention programs, intradistrict open enrollment, interdistrict open enrollment, voucher programs, tuition agreement programs, and charter schools[2] (Ysseldyke et al., 1992).

---

[1] Some parents have played a role in selecting the public school their children attend as part of their residential location decision (also described in the literature as the “school-housing connection”) or by enrolling their children into private schools. Thus, parental choice has always existed to some extent. Public school choice policies, however, extends this parental role.

[2] Magnet schools are schools with a particular pedagogical theme designed to attract minority and white students to attend the same school; postsecondary programs allow students to enroll in college courses and receive high school credit; dropout prevention programs allow eligible students that have not succeeded in a school and students at-risk to attend an alternative school; intradistrict open enrollment allows students to transfer to a different school within their district of residence; interdistrict open enrollment allows students to transfer to a school outside their district of residence; voucher programs allow students to attend a private school with tuition paid from public education funds; tuition agreement programs allow students to attend tuition-free any school of their choice if they are not assigned to a

Public school choice is advocated as a strategy to reform the public school system based on rational choice theory arguments that originate in Friedman (1955) and Chubb and Moe (1988)[3]. The principal argument asserts that, subjecting schools to market forces induces competition and forces the otherwise monopolized public school system to become more efficient. In addition, it improves the quality of education because the schools that fail to compete are likely to be forced out of the public school system. This, in turn, changes the incentives of schools to embrace change and provide high-quality programs in response to market pressures. Inevitably, school choice as a strategy for public school reform has generated much debate regarding issues of government involvement and students' and parents' rights for educational choice versus public education as a public good.

Advocates of public school choice attribute the lack of accountability, efficiency, and quality of public schools to the political governance of the education system. They argue that the administration of the public school system offers school districts little incentive for improvement (Chubb and Moe, 1990), arguing further that school choice promotes educational accountability (Young and Clinchy, 1992), equity and diversity (Young and Clinchy, 1992; Nelson et al., 1993), autonomy and competition that leads to general school system improvements (Cookson, 1994; Lieberman, 1990), encourages experimentation and risk taking, provides disadvantaged and minority students with better school options (Lieberman, 1990), etc.

---

school in their town; and charter schools are publicly-sponsored but autonomous schools that employ outcome-based education practices designed to enhance students' learning (Ysseldyke et al.,1992).

[3] Friedman was the first economist to call for an educational reform and proposed a voucher system that would allow parents to opt out of the public school system and enroll their children in private schools.

Critics question the appropriateness of a market-based reform because public education is a common good (Henig, 1994). Thus, they contest the implementation of public school choice in a market-like setting on grounds of equity and diversity. School choice may deepen the socioeconomic stratification in public schools as parents and students tend to separate themselves from those who are different[4] (Lutz, 1996). Parents may be uninformed to discern a school's academic quality in order to make choices that are in the best interest of their children (Cookson, 1994; Smith and Meier, 1995) or simply make choices based on convenience or other nonacademic factors such as religious beliefs or social class (Smith, 1995). Furthermore, schools that become financially distressed under competition for student enrollment are unlikely to provide equitable education to the remaining students. Thus, it is of concern that school choice may allocate the most disadvantaged students to the worst schools.

Despite the debate, the concept of school choice seems to appeal to the general public. A Gallup poll of 1987 reported that 71% of the public in the nation supported school choice (Smith, 1995). Results from a more recent poll – the 1997 Phi Delta Kappa/Gallup Poll of the Public Attitudes Towards the Public Schools – showed that although respondents were generally satisfied with their neighborhood school – 47% of the people surveyed gave their local school a grade of B or higher and 78% a grade of C or higher – 73% of the responders believed that school choice would improve the students' academic performance (Elam et al., 1997). This is also reflected in efforts to

[4] For this reason, an alternative to school choice that is currently being discussed (though not yet implemented) to address the stratification of the American school districts along the wealth and racial dimensions is to periodically redraw the school district boundaries (much like with the electoral districts) to ensure an economic and racial balance (e.g. Saiger, 2010).

pass some form of school choice legislation in every state and Washington D.C. (Cookson, 1994).

The empirical evidence from studies of various choice programs implemented throughout the U.S. regarding the main arguments brought forth by supporters and opponents of public school choice is debatable. Studies that examine the impact of school choice on educational outcomes – although generally do not report negative impacts – report effects that range from no change in educational outcomes to significant improvements in educational outcomes. Hoxby (2000, 2003) provides evidence that suggests that competition among public schools due to school choice substantially improves educational achievement, school productivity, and efficiency. Rapp (2000) analyses the influence of school choice on the behavior of school teachers. He is unable to establish a general influence of competitiveness on teacher effort, but his results suggest that one form of school choice, namely intradistrict enrollment, leads to more teacher effort. Belfield and Levin (2002) examined the evidence of over 41 empirical studies on the effects of competition on educational outcomes such as student test scores, graduation rates, educational expenditures, and teacher quality. They find that the majority of the studies are able to establish a link between competition (due to school choice) and educational outcomes and report beneficial effects of competition. However, the authors conclude that the magnitudes of these effects are only modest with respect to changes in the levels of competition. Belfield and Levin's (2002) analysis did not include empirical studies on voucher programs. However, the latter – Milwaukee's voucher program being one of the most extensively researched – also report effects anywhere from no gains in student

academic achievement to large gains (e.g. Witte et al., 1995; Green et al., 1999; Rouse, 1998). Similar effects are reported for open enrollment programs (e.g. O'Brien and Murdoch, 2000; Cullen et al., 2005; Carlson et al., 2011). Magnet schools seem to be more consistently associated with improved student outcomes (e.g. Metz, 1986; Crain et al., 1992).

The empirical evidence on the type of students that take advantage of choice programs is mixed as well. While some studies have found that children that attend choice programs are more able and advantaged suggesting that school choice increases sorting along the social class and ability dimension (e.g. Coleman et al., 1993; Witte, 1993; Buddin et al., 1998; Goldhaber et al., 1999), other studies find evidence of higher participation among minorities and low-income students (e.g. Duax, 1988; Lee et al., 1994; Schneider et al., 1996). The discrepancy in the effects of school choice programs on educational outcomes could be due to differences in policy design and the public school environment in which these policies are implemented or perhaps due to methodological differences in estimating competitive pressures and educational outcomes. The findings of these studies, however, point to the necessity of further analyzing different choice programs to obtain a better understanding of the operational aspects and the effectiveness of these market-based mechanisms.

The purpose of this study is to analyze the adoption decisions of school choice policies – interdistrict open enrollment being our policy of interest – from the standpoint of school districts. This form of educational choice is utilized in the U.S. by more than 40 states and serves to more students than any other type of school choice programs (Reback, 2008). In addition, unlike other school choice programs, it affects just about

every school district in a state. More specifically, we are interested in examining the determinants of adoption of interdistrict open enrollment policies by school districts within Ohio's public school system. Interdistrict open enrollment was first introduced in Ohio in 1989 with open enrollment options extended in 1997. The bill that passed in 1997 allowed Ohio's school districts to adopt one of following three policies: no open enrollment (i.e. school district chooses to accept only resident students), adjacent open enrollment (i.e. school district chooses to accept nonresident students only from adjacent districts), or statewide open enrollment (i.e. school district chooses to accept nonresident students from any district). Ohio's public school system operates on combined funding efforts from the state and school districts through local property taxes. Therefore, the implementation of interdistrict open enrollment policies carries financial ramifications for both the sending and receiving school districts. On the one hand, the sending school district loses funds because state funds allocated to a student follow the student across district lines. On the other hand, the receiving school district may need additional funds to educate nonresident students which are hard to obtain locally through increases in property taxes. Because the survival of schools and the quality of educational services they provide are directly linked to the school funding, school districts are faced with difficult decisions. Districts can choose not to participate, but any district can lose students to a nearby participating district.

Interdistrict open enrollment policies in Ohio's public schools have been examined in terms of operational outcomes such as administrative operability and technical effectiveness (e.g. Farrell 1994). Other studies have focused on whether or not there are quantitative differences in demographic characteristics, financial indicators, and

academic performance between school districts that decide to be open or closed to interdistrict enrollment (e.g. Metzler, 1996; Fowler, 1996; Crepage, 1999). In addition, these studies have collected descriptive evidence as to why districts decide to adopt one enrollment policy over another. Our study builds upon these studies to provide a quantitative analysis of the determinants of the adoption of interdistrict open enrollment policies by school districts and examine their relative importance in the probability of a school district to adopt any of the three interdistrict open enrollment alternatives. We estimate a spatially-explicit model of school districts' adoption decisions drawing from the literature on social interactions (e.g. Case, 1992; Besley and Case, 1995; Brett and Pinkse, 1997; Brueckner and Saavedra, 2001) that suggests that when faced with difficult decisions economic agents are influenced by decision makers in reference groups and either mimic their decisions or respond strategically to them. If this holds, we expect to find positive neighborhood influence on school districts' adoption decisions, an aspect of the school districts' adoption decisions that has been little explored. Rincke (2006) examined the neighborhood influence in the adoption of open enrollment policies using school district data from five states in the U.S. (Arkansas, California, Idaho, Massachusets, and Ohio) and found significant spatial effects suggesting that school districts' decisions are indeed heavily influenced by other school districts in reference groups. Our study provides further evidence that supports this view, but differs from this study in the following aspects. First, in Rincke's (2006) study the school district's adoption decision is binary; a school district decides to adopt a new policy, namely open enrollment, or not. Our set of policy alternatives includes all policy options provided by the interdistrict open enrollment statute. Second, because Rincke

(2006) was interested in neighborhood influence in diffusion of policy innovations in local jurisdictions, in his study spatial dependence is assumed to arise between school districts that belong to the same county. However, it may well be the case that school districts that are located in the periphery of a county may be influenced by nearby school districts that belong to an adjacent county. Our specification of the spatial weights matrix is more flexible and accommodates for this possibility.

The remainder of this chapter is organized as follows. The next section briefly describes the context, design, and implementation of interdistrict open enrollment policies in Ohio's public school system. The subsequent section describes our estimation framework followed by a section that describes the data set used to empirically estimate a spatially-explicit model of school districts' policy adoption decisions. The last section presents the estimation results and concludes.

Interdistrict Open Enrollment in Ohio

Following the recommendations of the State Board of Education, the Ohio Education 2000 Commission, and the Gillmor Commission, the Ohio legislature introduced school choice to Ohio's public school system through the Omnibus Educational Reform Act of 1989, Amended Senate Bill 140 (Ohio Department of Education, 1991). The Senate Bill 140 included three school choice policies, two of which were mandated – intradistrict open enrollment and post-secondary enrollment – and, the third option – interdistrict open enrollment – was voluntary. The intradistrict open enrollment option allowed students to enroll in a school within their residential district other than the neighborhood school assigned based on their residence. The post-secondary enrollment option allowed the 11th and 12th grade students to receive credit for post-secondary courses that would satisfy both high school graduation

requirements and college requirements. The third policy, interdistrict open enrollment, gave the school districts an option to accept nonresident students provided the students resided in adjacent districts. The interdistrict open enrollment statute set forth a time line for implementing the program and provided a list of procedures and requirements for applications and admissions to guide the school districts as well as parents and students that were interested in utilizing this option. The law prohibited school districts to select students for admission based on race, disability, academic performance, athletic talent, and disciplinary record; however, some provisions were put in place to prioritize enrollment (e.g. resident students or previously enrolled students over first-time applicants), maintain a racial balance, accommodate for districts' enrollment capacity, etc.

The implementation of interdistrict open enrollment started with a three-year pilot program in the school year 1990-1991and initially involved 3 (out of 611) school districts and 23 high school and elementary students. Participation of school districts and students in the remaining two years of the pilot program increased to 10 school districts and 115 students for the school year 1991-1992 and 49 school districts and 551 students for the school year 1992-1993 (Ohio Department of Education, 1993). Full implementation of the interdistrict open enrollment option was authorized for the fall of 1993. Prior to the beginning of the school year 1993-1994, the board of education of any school district had to adopt a resolution that either entirely prohibited the enrollment of students from adjacent districts or permitted the enrollment of students from all adjacent districts (Ohio Department of Education, 1991). Participation of school districts in the interdistrict open enrollment program increased to 301 in the school year 1993-

1994, 341 in the school year 1994-1995, 353 in the school year 1995-1996, and 376 in the school year 1996-1997 (Crepage, 1999).

The interdistrict open enrollment program was expanded with the passage of the Senate Bill 55 (SB 55) in July 1997, which allowed school districts to adopt on or after July 1998 one of the following policies that either: “entirely prohibits interdistrict open enrollment from any other school district (except for the students for whom tuition is paid); permits open enrollment of students from adjacent school districts, as under current law; or permits the open enrollment of students from any city, exempted village, or local school district that is not part of the joint vocational school district” (Crepage, 1999). SB 55 also expanded the post-secondary option to students in 9th and 10th grade. Participation of school districts in the expanded interdistrict enrollment program increased as well. Current participation for the school year 2010-2011 is 429 school districts (city, local, exempted village, joint vocational, and career centers) are open to statewide enrollment, 90 school districts are open to adjacent enrollment, and 144 school districts are closed (i.e. accepting only resident students) (Ohio Department of Education, 2010).

The implementation of Ohio’s interdistrict open enrollment program has raised many concerns over school funding. Ohio’s public school system is jointly funded by the state and school districts through local property taxes. Under the interdistrict open enrollment program, the amount of public funds allocated to a student’s education follows the student as he/she transfers across district lines. Therefore, the implementation of interdistrict open enrollment policies carries financial ramifications for both the sending and the receiving districts. Since the first year of the pilot program, a

formula has been used to transfer funds from the sending school district to the receiving school district[5]. In the 1997-1998 school year, Ohio introduced a “guarantee” figure to be used in the formula that calculates a school district’s basic state aid (Crepage, 1999). This figure represents the combined state and district funds allocated to a student for the school year adjusted by a county equalization factor (to allocate more funds to districts in high cost counties) and amounts to about 60-62% of the state average expenditure per student (Crepage, 1999). If a student enrolls in a school outside the residing district, the entire guarantee amount (computed based on the residing district) is deducted from the state aid the sending district receives for the school year and is added to the state aid of the receiving district. Because the educational cost per student differs across school districts, a receiving district does not always benefit from open enrollment. The amount of money that each participating student brings might not cover all educational costs, in which case additional funds need to be generated locally to educate nonresident students. Since parents of nonresident students are not part of the district’s tax base, it is difficult to justify increases in property taxes to obtain additional funds.

The implementation of the interdistrict open enrollment program has generated large variations in school funds for all school districts. Ruggles (1997) conducted a financial analysis of the interdistrict open enrollment program for the period starting in the 1993-1994 school year and ending in the 1996-1997 school year. His analysis shows variations in revenue gains and losses to Ohio’s school districts from

[5] For the school year 1990-1991, the school district that accepted a transfer student received $2,636 allocated to the student times ‘the cost-of-doing-business factor’ times the percent of the time the student attended, an amount that was deducted from the funds of the sending district (Ohio Department of Education, 1991).

participation (or lack thereof) in interdistrict open enrollment that range from a loss of about $7,333,000 for the Akron City School District to a gain of about $6,485,000 for the Coventry Local School District. These sizable changes in school funds suggest that any school district could take financial advantage of this option by improving their academic programs, facilities, implementing better marketing strategies, etc., – a main goal of the open enrollment program is to set up a more competitive atmosphere among school districts – however, it might create perverse financial incentives. Fowler (1996) finds that, all else equal, school leaders seem willing to compete for students and state funds, but under different conditions (e.g. when having to compete with wealthier districts) competition is less likely. Thus, the response of school districts to the policy needs to be evaluated within the financial and political constraints that each school district faces.

The literature that has examined the supply-side of Ohio's interdistrict open enrollment options have found systematic differences in characteristics between school districts that decide to adopt interdistrict open enrollment (open school districts), and those that forgo the option (closed school districts). Fowler (1996) examined the participation of Ohio's school districts in the interdistrict open enrollment option for the school year 1993-1994 – its first year of full implementation. The author reported that open districts tended to be those with low student enrollment, declining enrollment trends of previous five years, low minority enrollment (less that 1%), rural location, and below average per pupil expenditure ($3,501-$4,500). In contrast, closed districts tended to have high enrollment with increasing trends, suburban locations, minority enrollment between 11% and 20%, suburban locations, and above average per pupil expenditure (over $5,501). A study by Metzler (1996) for the school year 1994-1995

reported similar district differences. Open school districts had lower average daily membership, median family income, average class size, percentage of black and Asian students, percentage of students passing the proficiency tests, revenue and expenditure per pupil, average teacher salary and staff attendance, and higher percentage of economically and academically disadvantaged students. These findings were also corroborated by Crepage (1999) for the school districts that participated in the interdistrict open enrollment program for the school year 1996-1997.

The differences in the profiles of school districts that chose to adopt or forego interdistrict open enrollment options suggests that their adoption decisions are motivated by different factors. Fowler (1996) also surveyed school district leaders and gathered some descriptive evidence as to what were the major reasons that led them to the decision to adopt (or not) interdistrict open enrollment. The survey results indicate that the primary reasons for superintendents of open districts to choose open enrollment were concerns about losing student enrollment and state funds. Closed districts, on the other hand, chose to be closed mainly because of insufficient space (crowded buildings, large class sizes, maximum pupil-teacher ratio) and financial concern over subsidizing education costs of nonresident students with local funds. Interestingly, about a third of the superintendents of open districts indicated a major reason to be the influence from the decision of adjacent districts to be open, in some cases indicating that the decision was made collaboratively with adjacent districts. Similar evidence is provided in Crepage (1999).

Based on this evidence, we next model the school districts' policy adoption decisions in a manner that is consistent with their economic behavior, explicitly

accounting for spatial interdependence in decision making among neighboring school districts. We further note that, if spatial dependence is not accounted for, the policy adoption model is likely to produce misleading results.

Estimation Framework

We can start modeling the school districts' adoption decisions within a RUM framework by assuming that there are factors that jointly determine a school district's choice of a policy alternative. Some of these factors are observed by the researcher; for instance, characteristics of school districts such as student enrollment, class size, pupil-teacher ratio; financial factors such as revenues generated from property taxes and instruction expenses; and academic factors such student attendance, student performance, etc. Other factors are unobserved; for instance, philosophical beliefs or perceptions of school leaders regarding the different policy alternatives, etc. In addition to these factors, the school district's decision to adopt a particular policy alternative depends – as made evident in the preceding discussion – on the propensity of nearby school districts to choose that particular policy either strategically or collaboratively. Finally, it is reasonable to assume that the school district makes a decision to adopt a policy by choosing the policy alternative that generates the highest expected benefits. In this case, the term benefit needs to be interpreted within the context of the school district's objective for adopting the policy. For instance, if a school district is considering adopting an open enrollment policy because of low student enrollment, the optimal policy alternative would be the one that maximizes student enrollment, thus, the alternative actually adopted by the district and the choice observed by the researcher. Because some of the factors that determine the expected benefits a school district derives from various policy alternatives – which, in turn, determine a school district's

actual adoption of a policy alternative – are unobserved, we can set up a latent variable model of school districts' adoption decisions.

Consider a school district i that faces a decision to adopt one of L policy alternatives in the choice set and let l denote a general alternative. Let Y_{il}^* denote a latent variable representing the expected benefits that school district i generates from policy alternative l; Y_{il}^* depends, among other factors, on nearby school districts' expected benefits Y_{jl}^* as follows:

$$Y_{il}^* = Y\left(Y_{jl}^*, X_{il}\right) + \varepsilon_{il} \tag{5.1}$$

where X_{il} and ε_{il} represents school district i's observed and unobserved factors associated with policy alternative l, respectively. Then, under the model assumptions, the school district i will choose to adopt a particular policy alternative, say k, if and only if the policy alternative k generates the highest expected benefits when compared to all other policy alternatives:

$$Y_{ik}^* > Y_{il}^*, \forall k \neq l, l = \{1, \dots, L\} \tag{5.2}$$

Equivalently, the researcher observes the adoption of policy alternative k if:

$$d_{ik} = I(Y_{ik}^* > Y_{il}^*), \forall k \neq l \tag{5.3}$$

In this formulation, the policy adoption decision can be interpreted in the context of the propensity for adoption, thus, we can make probabilistic statements about the school district's adoption decision. The probability that a school district chooses policy alternative k over policy alternative l is the probability that the combination of observed and unobserved factors generates higher expected benefits under policy alternative k than under alternative l, resulting in that particular outcome. The probability that school district i adopts policy alternative k is given by:

$$P(d_{ik} = 1) = P(Y_{ik}^* > Y_{il}^*) = P\left[Y\left(Y_{jk}^*, X_{ik}\right) + \varepsilon_{ik} > Y\left(Y_{jl}^*, X_{il}\right) + \varepsilon_{il}\right] \quad (5.4)$$

Thus, the probability that a school district will adopt a particular policy alternative depends, among other things, on the underlying factors determining the probabilities of neighboring school districts to adopt the given policy alternative.

The choice probabilities for the policy adoption model in (5.4) can be empirically estimated once a functional form is specified for the expected benefits in (5.1) and a distributional assumption is made for the unobserved factors. The specification of the functional form of (5.1) requires specification of spatial interactions among school districts. We follow common practice and specify a spatial weights matrix W that quantifies the spatial relationships among neighboring school districts and estimate the following spatial autoregressive lag model:

$$Y_{ik}{}^* = \rho \sum_{j=1}^{N} w_{ij} Y_{jk}{}^* + X_i \beta_k + \varepsilon_{ik} \quad (5.5)$$

with, $d_{ik} = I(Y_{ik}^* > Y_{il}^*), \forall k \neq l$

where X denotes the model covariates; w_{ij} are the spatial weights assigned to school districts in locations i and j; and ρ and β_k are the parameters of interest. The parameter ρ measures the degree of neighborhood influence. A positive ρ means that a high propensity of adoption of a particular policy by a given school district positively affects the propensity of adoption of that alternative by a neighboring district, thus, increasing the probability of adoption of that policy alternative for both neighbors. Lastly, we assume that the error terms in (5.5) follow an IID type I extreme value distribution and estimate a spatial multinomial logit (SMNL) model using the estimation methodology developed in Chapter 3.

Data

To estimate the policy adoption model described in the previous section, we utilize data on Ohio's school districts for the school year 2000-2001. The dataset for this application is compiled from various sources and contains information on district demographic characteristics (Ohio Department of Education), income and racial heterogeneity indices (National Center for Education Statistics), property tax information (Ohio Department of Taxation), and geographic (latitude and longitude) information of location of the school districts[6].

Our sample is comprised of 609 school districts (city, local, and exempted village), of which, 365 are open to statewide enrollment, 90 are open to adjacent enrollment, and 154 have chosen not to participate in the open enrollment program (Table 5-1). The spatial distribution of the adoption of these enrollment policies is given by Figure 5-1. Figure 5-1 shows clustering of school districts that have adopted the same open enrollment policies, an indication of positive neighborhood influence in adoption decisions among neighboring districts. This influence seems to arise mainly between school districts in adjacent locations and spills over county lines. This visual evidence is useful for defining the spatial relationships between school districts (i.e. the type of the spatial weights matrix) to be used in the model. The districts that have chosen to be closed are primarily located in the suburbs of Cincinnati, Columbus, and Cleveland metropolitan areas. These areas are characterized by high student enrollment and higher than average median income. In contrast, school districts that are open are

[6] The dataset for this study was compiled and made available by Dr. David M. Brasington, University of Cincinnati.

located mainly in rural areas and urban areas that are characterized by low student population and lower than average median income (Figure 5-2).

A description of the variables used in the model along with descriptive statistics is provided in Table 5-2 and Table 5-3. The dependent variable in our model is an indicator variable for the school districts' open enrollment status (Openenroll), with three policy alternatives: 1) statewide open enrollment, 2) adjacent open enrollment, and 3) no open enrollment. To explain the school districts' adoption decisions we include the following covariates. First, we include the fall school district enrollment for the school year 2000-2001 (Falladm00), computed as kindergarten through grade 12 enrollment adjusted by adding non-attending pupils and subtracting any unauthorized attendance. The mean student enrollment for the school year 2000-2001 is about 2,850 pupils. Because many school districts that have decided to adopt interdistrict open enrollment are characterized by low minority enrollment, we include the percentage of student enrollment that is non-white (Pctmin00). The average percentage of minority student enrollment in our sample is 7.2%. Other important demographic characteristics that likely affect the school districts' adoption decisions are income and racial heterogeneity, so we use racial heterogeneity (Racehet00) and income heterogeneity (Inchet00) measured by a Leik (1966) index for the school district in the school year 2000-2001. A typical school district in our sample is characterized by relatively low racial heterogeneity but relatively high income heterogeneity. We are also interested in whether changes in these measures affect the choice probability of a particular policy alternative, so the change in racial heterogeneity (Racehet_ch9000) and income heterogeneity (Inchet_ch9000) in the school district between 1990 and 2000 are also

included. The average change in the racial heterogeneity index is positive indicating an increase in racial heterogeneity from the previous decade for the typical school district. In contrast, the average change in income heterogeneity is negative indicating a reduction in income heterogeneity. The decision of a school district to be open or closed to interdistrict open enrollment also largely depends on school district enrollment capacity. Among various measures of capacity (e.g. class size, building capacity, student-teacher ratio), we use the change in the student-teacher ratio in the school district from the previous decade (Tadmprtc_ch9000) to investigate whether or not a change in capacity strongly effects the probability of adoption of any of the three policy alternatives. The average change in the student-teacher ratio is negative showing a decrease in the number of students assigned to a teacher compared to the previous decade. The percentage of students in each school district proficient and above in math based on a 9th grade proficiency test (Math00) is included as a measure of academic performance.

Another important aspect of the school districts' open enrollment adoption decisions is the financial aspect. School districts may choose to be open or closed depending on whether or not the state funds that accompany the transfer students cover all the educational expenses. Thus, the school districts' decisions will highly depend on current availability of local funds as well as their ability to generate additional educational funds locally. For this reason, we include the local school revenues per pupil for the school year 2000-2001(Localxpa00) and the change in these revenues from the previous decade (Localxpa_ch9000). Local school revenues are obtained by dividing a school district's Class 1 (agricultural and residential) tax collections by the fall

school district enrollment. For the school year 2000-2001 the average per pupil revenue generated locally through property taxes is about $2,000. This figure has increased by an average of $890 from the previous decade. In addition, we include the Class 1 property tax rate in effective mills (Ptaxrate00), the percentage of real property in the school district that is of agricultural value (Pctag00), the change in the percent of real property in the school district that is of agricultural value (Pctag_ch9000), and Class 2 property value per pupil (Cl2valpera00) to examine the relative importance of the factors that determine the school district's ability to generate local funds in the probability of adoption of the three policy alternatives.

Finally, to carry out our analysis we need to define the spatial relationships between the school districts and specify a spatial weights matrix W. Because of the heterogeneity in the size of the different school districts (Figure 5-1), specifying a distance-based spatial weights matrix would result in a large number of neighbors for small districts and few (or perhaps no) neighbors for large districts (Anselin, 2002). So we specify the spatial weights matrix based on a fixed number of nearest neighbors to ensure there will be some neighbors for every school district. A school district in Ohio has on average 6 nearest neighbors; however, the number of neighbors varies from 1 (or zero for the few island districts) to 15 first-order neighbors. Using the school districts' geographic information we specify four spatial weights matrices based on 3, 6, 8, and 10 nearest neighbors. Note that the specification of the spatial weights matrix based on 3 nearest neighbors, while it may capture well the neighborhood structure for the larger districts, it is likely to misrepresent the neighborhood for the moderate to small size districts. On the other hand, the specification of the spatial weights matrix

based on 10 nearest neighbors is likely to misrepresent the neighborhood for the large districts as it will include second-order, third-order or even higher order neighbors between which little to no spatial interaction is expected to be present. The remaining two specifications (based on 6 and 8 nearest neighbors) are more likely to capture the neighborhood structure of the typical school district. Following common practice, all matrices set $w_{ii} = 0$ and are row-standardized (rows sum to one).

Estimation Results

The results for the different specifications of the SMNL model (based on the different specifications of W) along with MNL estimates (as benchmark estimates) are presented in Table 5-4. The base category chosen for the analysis is the statewide open enrollment policy. In general, the estimates from these models align with expectations. In addition, the SMNL estimates are robust across the different spatial specifications; occasional changes in sign occur only for estimates that are not statistically different from zero.

Because the MNL and SMNL models are nonlinear in parameters, the estimated coefficients only indicate the direction of the effect of each hypothesized factor in the relative probability of adoption of adjacent open enrollment and no open enrollment alternatives (relative to statewide open enrollment). Marginal effects, on the other hand, allow interpretation of the magnitude of each factor in terms of its importance in the probability of adoption of the three policy alternatives. Because the spatial effects in our spatial model specifications are global – a change in a covariate for a given school distinct will affect not only the probability of policy adoption for the own district, but also the policy adoption probability for the remaining school districts – the marginal effects

are not point estimates, but $N \times N$ matrices with the diagonal elements representing the direct (own) effects of a change in the explanatory variable and the off-diagonal elements representing the indirect (spillover) effects.

The marginal effects for our SMNL model are given by:

$$ME_{ik} = P_{ik}(\beta_k - \sum_l P_{il}\,\beta_l) \odot (I - \rho W)^{-1} \odot (\iota_i \frac{1}{\sigma_i}) \tag{5.6}$$

where ι_N denotes an $Nx1$ vector of ones and $\odot$ represents element-by-element (Hadamard) multiplication. To provide scalar measures of the marginal effects, we follow LeSage and Pace's (2009) suggestion and average over the diagonal elements to obtain a measure of the direct effects, average the row sums to produce a measure of the total effect, and take the difference between the two measures to get a measure of the indirect effects as follows:

$$\begin{aligned} \overline{ME}\,(k)_{direct} &= N^{-1}\,tr(ME_k) \\ \overline{ME}\,(k)_{total} &= N^{-1}\,\iota_N'(ME_k)\iota_N \\ \overline{ME}(k)_{indirect} &= \overline{ME}\,(k)_{total} - \overline{ME}\,(k)_{direct} \end{aligned} \tag{5.7}$$

The estimates of the marginal effects for the MNL model and the different specifications of the SMNL model are presented in Table 5-5. These estimates are also fairly robust across the different specifications of the SMNL model as well as across the spatial and non-spatial specifications. A marginal effect produced by the MNL model measures the change in the probability of adoption of a given policy alternative associated with a change in the average observation (school district) for a particular explanatory variable. The interpretation of the marginal effects produced by a spatial discrete-choice model is analogous with the exception that a change in a particular explanatory variable for a given school district generates multiple effects. The direct

effect measures the impact of a one-unit change in a covariate on the probability of adoption of a given policy alternative for the school district of interest. The indirect effect measures the impact of a one-unit change in a covariate on the probability of adoption of a given policy alternative for neighboring school districts. The total effect, as the sum of the preceding two effects, shows the average total impact of a change in a covariate for a given school district on the probability of adoption of a given policy alternative for the school district of interest. For comparison purposes, a marginal effect produced by the MNL model for a specific policy alternative would be interpreted as a direct effect since the indirect effects in non-spatial models are zero (LeSage and Pace, 2009). Prior to interpreting the estimates of the marginal effects, it is worth remarking that, for the majority of the explanatory variables, the magnitudes of the marginal effects for the MNL model are comparable to the magnitudes of the direct effects produced by the SMNL model specifications. However, when we look at the total effects, the latter are much larger due to the indirect effects, pointing to the importance of the neighborhood influence on school districts' policy adoption decisions, a decision making aspect that the MNL model fails to capture. We note further, that across all specifications of the SMNL model, the direct effects and indirect effects have the same sign, a result that is consistent with the presence of positive neighborhood influence.

Starting the analysis with the school districts' demographic characteristics, the estimated coefficients for fall student enrollment (Falladm00) are not statistically significant, thus there is uncertainty about the direction of the effect of this factor on the probability of adoption of adjacent and no open enrollment alternatives (relative to statewide enrollment) possibly due to being fairly contemporaneous. As expected, the

estimates of the marginal effects with respect to this factor are also imprecisely estimated in both the MNL and SMNL specifications. The effect of the percentage of minority student enrollment (Pctmin00) is statistically significant in all model specifications. An increase in the percentage of minority enrollment is associated with a decrease in the probability of adoption of statewide open enrollment and an increase in the probability of adoption of the other two alternatives. This is expected since statewide open enrollment is typically adopted in school districts with low minority population. The Leik (1966) index measure of racial heterogeneity (Racehet00) negatively affects the adoption probability of adjacent open enrollment and no open enrollment relative to statewide open enrollment. We would expect that an increase in racial heterogeneity be associated with a decrease in the adoption probability of statewide open enrollment since the adoption of this policy alternative occurs in school districts that have a rather homogeneous racial composition. Our expectations are confirmed by the effects of the change in the racial index from the previous decade. A positive change in the racial heterogeneity index from the previous decade (Racehet_ch9000) – denoting an increase in racial heterogeneity – negatively affects the probability of adoption of statewide enrollment and increases the adoption probability of the no open enrollment alternative. Similar effects are found for the income heterogeneity index (Inchet00). An increase in income heterogeneity increases the probability of adoption of statewide open enrollment and decreases the adoption probability of the no open enrollment policy. The effect of income heterogeneity in the probability of adoption of the adjacent open enrollment alternative and the effects of the change in income heterogeneity from the previous decade (Inchet_ch9000) are unclear since their estimates are not

statistically significant. The effect of the capacity variable – the change in student-teacher ratio (Tadmprtc_ch9000) – is positive for the adoption probability of adjacent enrollment and negative for the statewide open enrollment alternative. The average change in the student-teacher ratio from the previous school decade is negative, indicating an average decrease in the number of students per teacher. This could explain the positive effect in the probability of adoption of adjacent open enrollment as more students can be accommodated if school districts have not reached the maximum student-teacher ratio. The academic student performance – measured by the percentage of students in each school district proficient and above in math (Math00) – increases the probability of a school district choosing not to participate in open enrollment while decreases the probability of adoption of the statewide enrollment policy. School districts that choose to be closed are generally characterized by higher student performance (higher percentage of students that pass proficiency tests) than open enrollment school districts.

Regarding the effects of the financial factors, the local school revenue per pupil (Localxpa00) is expected to positively affect the probability of a school district not participating in open enrollment relative to statewide enrollment; however, this is only confirmed by the MNL marginal effect estimate of the no open enrollment alternative. Local revenues are especially important to those school districts with above average per student expenses, which are typically the school districts that do not participate in the open enrollment options. However, when school districts are able to generate more funds locally they are more likely to participate in open enrollment as shown by (Localxpa_ch9000). This is also a result confirmed by the MNL marginal effect estimate

of the no open enrollment policy alternative. The effect of the Class 1 property tax rate (Ptaxrate00) is uncertain because the marginal effect estimates across all model specifications are not statistically different from zero. The amount of property with agricultural value (Pctag00) has a negative effect in the probability of a school district to be closed but a positive effect on the probability of adoption of adjacent open enrollment. This is expected since the school districts that are open to interdistrict open enrollment are mainly located in rural areas. The effect of the Class 2 property value per pupil in the probability of adoption of open enrollment policies is negative for the MNL model, but this effect becomes statistically insignificant in the specifications of the spatial models.

Most importantly, the estimates of the spatial lag parameter (ρ) produced by the SMNL model vary from 0.32 to 0.63, depending on the specification of W. The estimate of the spatial lag parameter shows weak presence of spatial dependence for the specification of W based on 3 nearest neighbors – likely indicating that the spatial weights matrix based on 3 nearest neighbors does not represent well the range of interactions among school districts – but this dependence becomes stronger as the neighborhood size increases. As a result, the indirect effects obtained for the spatial model specification based on 3 nearest neighbors are generally not statistically significant. As anticipated, the neighborhood influence is strongest for the neighborhood specification based on 8 nearest neighbors. This influence decreases as the size of the neighborhood becomes larger to include 10 nearest neighbors. Interestingly, the indirect effects of some of the school districts decision factors such as the percentage of minority enrollment, income heterogeneity, and academic performance (based on the

math proficiency tests) are larger than the respective direct effects for the last two spatial model specifications, indicating an especially strong mutual influence of neighboring school districts along these enrollment dimensions.

These results confirm our expectations that there is positive neighborhood influence on school districts' adoption decisions suggesting that school districts' decisions are influenced by other school districts with similar demographic characteristics and financial constraints (i.e. school districts in reference groups). Thus, a high propensity of a given school district to adopt a particular open enrollment policy alternative positively affects the propensity of a nearby school district to adopt the given alternative, increasing the probability that the two neighboring districts will adopt the same policy. This, in turn, implies clustering of school districts with similar enrollment policies, which complies with the spatial pattern of open enrollment policies observed in Figure 5-1.

Finally, the NxN matrices of marginal effects obtained by the SMNL model for each of the factors that affects school districts' enrollment decisions provide information not only regarding the direct impacts of each of these factors on the probability of adoption of open enrollment policy alternatives, but also the indirect impacts that spill over from (or to) the surrounding school districts. As it was made evident in the previous chapter, the scalar measures reported in the section are only one way to summarize and analyze these effects. The NxN matrices of marginal effects can be partitioned to represent different neighborhoods of interest and the marginal effects can be analyzed similarity. The latter is particularly helpful to identify the locations with the most influence (i.e. the reference groups).

Final Remarks

This chapter provides a quantitative analysis of the determinants of the adoption of interdistrict open enrollment policies by school districts, particularly focusing on the neighborhood influence in school districts' policy adoption decisions. It estimates a spatially-explicit model of policy adoption decisions of Ohio's school districts to examine the importance of these determinants in the adoption probability of three interdistrict open enrollment alternatives; statewide open enrollment, adjacent open enrollment, and no open enrollment.

The empirical results substantiate extant descriptive evidence regarding the determinants of adoption of open enrollment policies. Among the most influential factors are minority population, racial and income heterogeneity, changes in student-teacher ratio, student academic performance, and financial factors that indicate a school district's current availability of local funds as well as its ability to generate additional funds locally. More importantly, the results show strong neighborhood influence in policy adoption decisions.

The results from our analysis have the following empirical and policy implications. Empirically, a policy adoption model that ignores the potential neighborhood influence is likely to suffer from specification error and produce potentially misleading results. From a policy viewpoint, the presence of neighborhood effects provides further support to the view that public policy adoption is stimulated by interaction among decision makers at a local level. Thus, the adoption of new public policies much like the diffusion of technology innovations is largely influenced by reference groups. A spatial analysis of these neighborhood effects can help identify the most influential locations (i.e. reference groups).

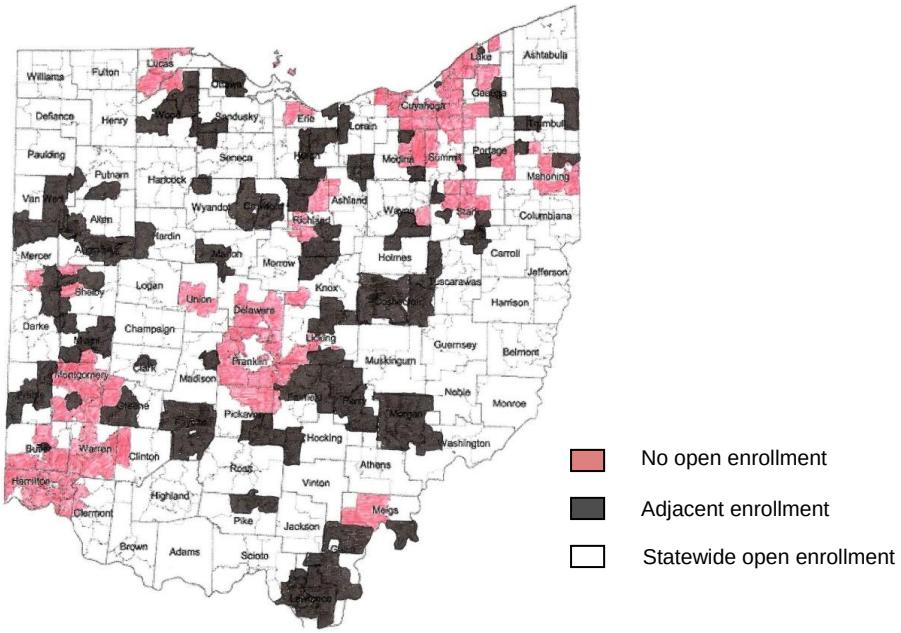
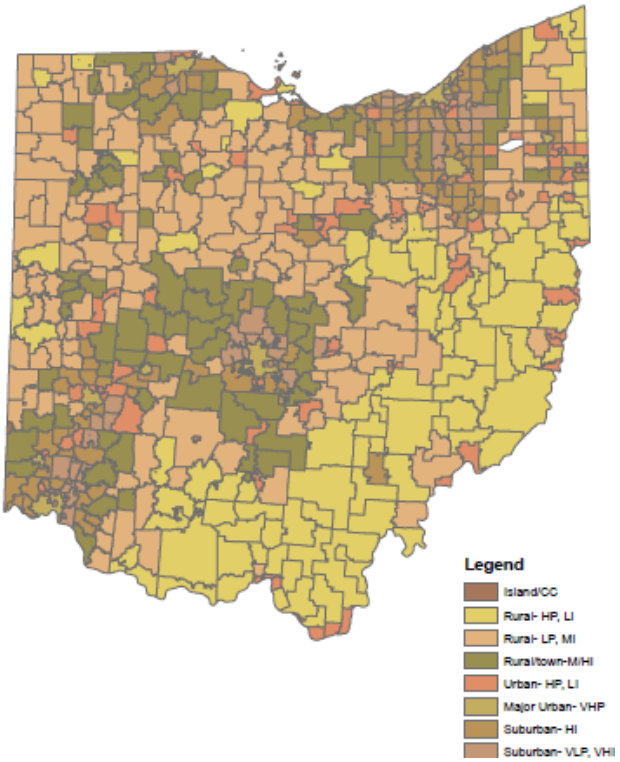


Figure 5-1. Interdistrict open enrollment policies in Ohio



Source: Ohio Department of Education

Figure 5-2. Typology of Ohio's school districts

Table 5-1. School districts in each enrollment category

Open Enrollment Policy	School Districts	Share (%)
Statewide	365.00	59.93
Adjacent	90.00	14.78
No open	154.00	25.29
Total	609.00	100.00

Table 5-2. Policy adoption model: variable description

Variable	Description
Openenroll	Indicator variable for school district's open enrollment status (1-statewide enrollment, 2-adjacent enrollment, 3-no open enrollment)
Falladm00	Fall school district enrollment (in thousands of students) for school year 2000-2001
Pctmin00	Percent of student enrollment that is non-white for school year 2000-2001
Tadmprtc_ch9000	Change in student-teacher ratio in school district between 1990 and 2000
Ptaxrate00	Class 1 property tax rate in effective mills in school district for school year 2000-2001
Pctag00	Percent of real property in school district that is agricultural value for 2000
Pctag_ch9000	Change in percent of real property in school district that is agricultural value (1990-2000)
Racehet00	Racial heterogeneity in school district in 2000, measured by Leik (1966) index
Racehet_ch9000	Change in racial heterogeneity in school district between 1990 and 2000
Inchet00	Income heterogeneity in school district in 2000, measured by Leik (1966) index
Inchet_ch9000	Change in income heterogeneity in school district between 1990 and 2000
Localxpa00	Local school revenues per pupil (in thousands of dollars) for school year 2000-2001
Localxpa_ch9000	Change in local school revenues per pupil between 1990 and 2000
Math00	Percent of students in each school district proficient or above in math section of the school year 2000-2001 9th grade proficiency test
Cl2valpera00	Class 2 property value per pupil (in thousands of dollars)

Table 5-3. Policy adoption model: variable descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
openenroll	1.65	0.86	1.00	3.00
falladm00	2.85	4.98	0.04	72.28
pctmin00	7.23	13.77	0.00	100.00
tadmprtc_ch9000	-0.25	1.98	-10.20	8.10
ptaxrate00	26.88	5.23	18.35	61.93
pctag00	13.04	12.79	0.00	56.07
pctag_ch9000	-2.50	3.05	-14.30	3.63
racehet00	0.06	0.05	0.00	0.45
racehet_ch9000	0.02	0.03	-0.15	0.18
inchet00	0.71	0.10	0.49	1.21
inchet_ch9000	-0.35	0.10	-0.72	0.09
localxpa00	2.02	1.43	0.36	19.42
localxpa_ch9000	0.89	0.66	-0.55	7.11
math00	75.22	11.38	28.50	98.70
cl2valpera00	17.55	21.71	0.50	232.31

Table 5-4. Estimated coefficients of the policy adoption model

			SMNL Models							
	MNL Model		W_3nn		W_6nn		W_8nn		W_10nn	
	Est.	St. Err.	Est.	Est.	St. Err.	St. Err.	St. Err.	St. Err.	St. Err.	St. Err.
falladm00:										
βadj	-0.058	0.072	-0.051	0.170	-0.066	0.174	-0.029	0.177	-0.109	0.180
βno	-0.002	0.029	-0.014	0.029	-0.028	0.031	-0.034	0.031	-0.025	0.030
pctmin00:										
βadj	0.064	0.024	0.073	0.027	0.078	0.027	0.072	0.027	0.080	0.027
βno	0.091	0.025	0.103	0.030	0.109	0.031	0.104	0.030	0.105	0.031
racehet00:										
βadj	-20.503	7.238	-23.735	11.890	-25.303	11.878	-22.823	11.743	-24.878	11.846
βno	-17.426	7.217	-23.868	11.017	-23.974	11.495	-22.658	11.157	-23.937	11.101
racehet_ch9000:										
βadj	16.875	8.125	19.328	12.314	22.160	12.203	19.196	12.155	21.298	12.267
βno	15.874	8.467	18.820	12.213	15.540	12.914	14.439	12.548	18.456	12.656
inchet00:										
βadj	-2.538	1.656	-2.076	1.887	-2.003	1.864	-2.296	1.858	-2.362	1.873
βno	-11.190	2.413	-11.978	4.005	-11.538	4.060	-10.704	4.100	-9.976	4.103
inchet_ch9000:										
βadj	0.786	1.499	0.637	1.677	0.296	1.656	0.895	1.660	0.788	1.669
βno	1.767	1.841	1.825	2.457	1.167	2.506	0.799	2.528	0.872	2.536
tadmprtc_ch9000:										
βadj	0.145	0.064	0.136	0.069	0.122	0.070	0.146	0.070	0.145	0.070
βno	0.023	0.079	0.119	0.102	0.076	0.101	0.085	0.101	0.071	0.102
math00:										
βadj	0.010	0.015	0.008	0.017	0.003	0.017	0.012	0.018	0.009	0.018
βno	0.063	0.020	0.081	0.031	0.090	0.031	0.094	0.031	0.086	0.031
localxpa00:										
βadj	-0.483	0.500	-0.096	1.024	0.680	0.966	-0.148	0.971	-0.040	0.997
βno	1.601	0.396	0.850	0.846	0.533	0.808	0.191	0.849	0.525	0.904

Table 5-4. Continued

			SMNL Models							
	MNL Model		W_3nn		W_6nn		W_8nn		W_10nn	
	Est.	St. Err.	Est.	Est.	St. Err.	St. Err.	St. Err.	St. Err.	St. Err.	St. Err.
localxpa_ch9000:										
βadj	0.664	0.773	0.424	1.401	-0.581	1.346	0.426	1.335	0.210	1.345
βno	-1.766	0.613	-1.056	1.025	-0.659	0.988	-0.322	1.034	-0.661	1.084
ptaxrate00:										
βadj	0.052	0.041	0.043	0.051	0.021	0.051	0.031	0.052	0.031	0.053
βno	-0.001	0.040	0.007	0.049	0.003	0.049	0.004	0.048	0.005	0.049
pctag00:										
βadj	0.016	0.013	0.005	0.017	0.014	0.017	0.014	0.017	0.014	0.017
βno	-0.093	0.026	-0.134	0.054	-0.164	0.053	-0.160	0.052	-0.137	0.052
pctag_ch9000:										
βadj	0.090	0.047	0.093	0.048	0.075	0.049	0.090	0.048	0.096	0.048
βno	-0.023	0.076	-0.072	0.119	-0.122	0.116	-0.103	0.117	-0.083	0.117
cl2valpera00:										
βadj	-0.022	0.019	-0.053	0.043	-0.011	0.042	-0.027	0.042	-0.017	0.043
βno	0.031	0.013	0.028	0.023	0.017	0.022	0.011	0.022	0.019	0.023
rho	--	--	0.322	0.188	0.418	0.217	0.626	0.247	0.524	0.253

Note: The columns for the SMNL model correspond to different specifications of the spatial weights matrix W that vary the number of nearest neighbors to 3, 6, 8, and 10, respectively. See text for details.

Table 5-5. Marginal effects for the estimated coefficients of the policy adoption model

		SMNL Model (W_3nn)			SMNL Model (W_6nn)			SMNL Model (W_8nn)			SMNL Model (W_10nn)		
	MNL Model	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
falladm00:													
stw	0.0062	0.0048	0.0020	0.0068	0.0064	0.0041	0.0105	0.0030	0.0046	0.0076	0.0075	0.0074	0.0149
	(0.0083)	(0.0135)	(0.0057)	(0.0190)	(0.0133)	(0.0089)	(0.0218)	(0.0080)	(0.0117)	(0.0193)	(0.0109)	(0.0109)	(0.0205)
adj	-0.0069	-0.0044	-0.0018	-0.0063	-0.0056	-0.0036	-0.0091	-0.0013	-0.0017	-0.0030	-0.0067	-0.0066	-0.0133
	(0.0085)	(0.0152)	(0.0063)	(0.0214)	(0.0160)	(0.0106)	(0.0264)	(0.0087)	(0.0122)	(0.0209)	(0.0121)	(0.0115)	(0.0227)
no	0.0007	-0.0004	-0.0002	-0.0005	-0.0008	-0.0006	-0.0014	-0.0017	-0.0029	-0.0046	-0.0008	-0.0009	-0.0016
	(0.0031)	(0.0027)	(0.0013)	(0.0039)	(0.0035)	(0.0027)	(0.0062)	(0.0018)	(0.0045)	(0.0060)	(0.0022)	(0.0029)	(0.0050)
pctmin00:													
stw	-0.0148	-0.0116	-0.0050	-0.0166	-0.0114	-0.0077	-0.0191	-0.0083	-0.0131	-0.0213	-0.0100	-0.0106	-0.0205
	(0.0041)	(0.0030)	(0.0035)	(0.0039)	(0.0033)	(0.0023)	(0.0048)	(0.0041)	(0.0052)	(0.0083)	(0.0041)	(0.0051)	(0.0083)
adj	0.0061	0.0054	0.0022	0.0076	0.0054	0.0033	0.0087	0.0030	0.0041	0.0072	0.0044	0.0042	0.0086
	(0.0027)	(0.0025)	(0.0016)	(0.0033)	(0.0025)	(0.0015)	(0.0041)	(0.0019)	(0.0021)	(0.0040)	(0.0031)	(0.0024)	(0.0037)
no	0.0087	0.0061	0.0029	0.0090	0.0060	0.0044	0.0104	0.0052	0.0089	0.0142	0.0056	0.0064	0.0119
	(0.0026)	(0.0019)	(0.0022)	(0.0031)	(0.0021)	(0.0024)	(0.0040)	(0.0020)	(0.0044)	(0.0064)	(0.0019)	(0.0034)	(0.0053)
racehet00:													
stw	3.6687	3.2156	1.3803	4.5959	3.1265	2.0826	5.2091	2.1204	3.2901	5.4104	2.6475	2.7693	5.4167
	(1.1441)	(1.2216)	(0.9573)	(1.5407)	(1.2892)	(0.7163)	(1.7588)	(1.3834)	(1.9837)	(2.0009)	(1.1111)	(1.3201)	(1.8327)
adj	-2.1497	-1.8645	-0.7500	-2.6146	-1.9332	-1.2180	-3.1513	-1.0160	-1.4004	-2.4164	-1.4213	-1.3689	-2.7902
	(0.8445)	(1.1112)	(0.5975)	(1.4529)	(1.2228)	(0.9971)	(1.8359)	(1.0646)	(0.9084)	(1.6875)	(0.1538)	(0.0907)	(0.1595)
no	-1.5191	-1.3511	-0.6303	-1.9813	-1.1932	-0.8646	-2.0578	-1.1044	-1.8897	-2.9940	-1.2262	-1.4003	-2.6266
	(0.7646)	(0.7525)	(0.5392)	(1.0956)	(0.8370)	(0.8013)	(1.4011)	(0.6709)	(1.7727)	(1.9941)	(0.5948)	(1.3209)	(1.6090)
racehet_ch9000:													
stw	-3.1575	-2.5836	-1.1077	-3.6913	-2.4616	-1.6235	-4.0852	-1.5581	-2.3817	-3.9399	-2.1617	-2.2510	-4.4127
	(1.2940)	(1.2541)	(0.8381)	(1.6666)	(1.1936)	(1.1806)	(1.8544)	(1.2636)	(1.6182)	(2.1286)	(1.0365)	(1.1943)	(1.9697)
adj	1.7434	1.5248	0.6137	2.1386	1.7823	1.1320	2.9143	0.8762	1.2150	2.0913	1.2296	1.1866	-4.4127
	(0.9534)	(1.1313)	(0.5665)	(1.5277)	(1.2552)	(1.0264)	(1.9681)	(0.9784)	(0.9413)	(1.6886)	(1.0815)	(0.9049)	(1.6290)
no	1.4141	1.0588	0.4940	1.5528	0.6794	0.4916	1.1709	0.6819	1.1667	1.8486	0.9321	1.0644	1.9965
	(0.9155)	(0.8475)	(0.5041)	(1.2321)	(0.9286)	(0.6965)	(1.5458)	(0.7122)	(1.4156)	(1.9243)	(0.7727)	(1.1708)	(1.6905)

Table 5-5. Continued

		SMNL Model (W_3nn)			SMNL Model (W_6nn)			SMNL Model (W_8nn)			SMNL Model (W_10nn)		
	MNL Model	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
inchet00:													
stw	1.2685	0.8405	0.3816	1.2222	0.7361	0.5189	1.2550	0.6242	1.0340	1.6583	0.6479	0.7157	1.3636
	(0.2919)	(0.2523)	(0.2775)	(0.3665)	(0.2683)	(0.1703)	(0.4139)	(0.2933)	(0.5280)	(0.6228)	(0.2922)	(0.3067)	(0.4543)
adj	-0.1139	-0.0579	-0.0177	-0.0757	0.0045	0.0189	0.0234	-0.0625	-0.0730	-0.1355	-0.0864	-0.0743	-0.1608
	(0.1960)	(0.1786)	(0.0750)	(0.2525)	(0.2373)	(0.1627)	(0.3995)	(0.1024)	(0.1404)	(0.2332)	(0.1256)	(0.1297)	(0.2434)
no	-1.1546	-0.7826	-0.3639	-1.1465	-0.7406	-0.5378	-1.2784	-0.5617	-0.9610	-1.5227	-0.5614	-0.6414	-1.2028
	(0.2567)	(0.2316)	(0.2615)	(0.3347)	(0.2639)	(0.1916)	(0.4361)	(0.2579)	(0.4032)	(0.6233)	(0.2587)	(0.2616)	(0.4270)
inchet_ch9000:													
stw	-0.2397	-0.1532	-0.0683	-0.2215	-0.0815	-0.0569	-0.1384	-0.0788	-0.1216	-0.2003	-0.0897	-0.0944	-0.1842
	(0.2405)	(0.1942)	(0.0858)	(0.2680)	(0.1818)	(0.1161)	(0.2924)	(0.1602)	(0.2172)	(0.3659)	(0.1739)	(0.1618)	(0.3246)
adj	0.0639	0.0374	0.0144	0.0518	0.0081	0.0036	0.0118	0.0403	0.0556	0.0959	0.0443	0.0426	0.0869
	(0.1770)	(0.1519)	(0.0634)	(0.2145)	(0.1591)	(0.1044)	(0.2634)	(0.0863)	(0.1171)	(0.1981)	(0.1072)	(0.1066)	(0.2096)
no	0.1758	0.1158	0.0539	0.1697	0.0734	0.0533	0.1267	0.0385	0.0659	0.1044	0.0454	0.0519	0.0973
	(0.1956)	(0.1666)	(0.0744)	(0.2326)	(0.1720)	(0.1140)	(0.2812)	(0.1373)	(0.2135)	(0.3473)	(0.1489)	(0.1533)	(0.2987)
tadmprtc_ch9000:													
stw	-0.0169	-0.0174	-0.0074	-0.0249	-0.0130	-0.0086	-0.0216	-0.0106	-0.0160	-0.0267	-0.0119	-0.0121	-0.0240
	(0.0103)	(0.0077)	(0.0067)	(0.0123)	(0.0073)	(0.0079)	(0.0128)	(0.0073)	(0.0134)	0.015896	(0.0073)	(0.0108)	(0.0140)
adj	0.0169	0.0109	0.0044	0.0153	0.0100	0.0063	0.0163	0.0068	0.0095	0.0162	0.0087	0.0084	0.0171
	(0.0075)	(0.0064)	(0.0035)	(0.0084)	(0.0066)	(0.0055)	(0.0103)	(0.0035)	(0.0040)	(0.0073)	(0.0036)	(0.0031)	(0.0076)
no	0.0000	0.0065	0.0031	0.0096	0.0031	0.0022	0.0053	0.0038	0.0066	0.0104	0.0032	0.0037	0.0069
	(0.0083)	(0.0070)	(0.0046)	(0.0111)	(0.0068)	(0.0056)	(0.0122)	(0.0054)	(0.0124)	(0.0173)	(0.0059)	(0.0084)	(0.0140)
math00:													
stw	-0.0067	-0.0052	-0.0024	-0.0076	-0.0048	-0.0034	-0.0082	-0.0051	-0.0086	-0.0137	-0.0049	-0.0055	-0.0105
	(0.0025)	(0.0021)	(0.0023)	(0.0038)	(0.0021)	(0.0033)	(0.0047)	(0.0018)	(0.0042)	(0.0065)	(0.0018)	(0.0057)	(0.0063)
adj	0.0001	-0.0001	-0.0001	-0.0002	-0.0012	-0.0009	-0.0021	0.0002	0.0001	0.0002	-0.0000	-0.0001	-0.0001
	(0.0018)	(0.0017)	(0.0007)	(0.0023)	(0.0024)	(0.0018)	(0.0041)	(0.0010)	(0.0014)	(0.0024)	(0.0013)	(0.0013)	(0.0025)
no	0.0066	0.0053	0.0025	0.0078	0.0060	0.0043	0.0103	0.0050	0.0085	0.0134	0.0049	0.0056	0.0106
	(0.0020)	(0.0019)	(0.0023)	(0.0037)	(0.0022)	(0.0022)	(0.0045)	(0.0017)	(0.0043)	(0.0075)	(0.0026)	(0.0026)	(0.0061)

Table 5-5. Continued

		SMNL Model (W_3nn)			SMNL Model (W_6nn)			SMNL Model (W_8nn)			SMNL Model (W_10nn)		
	MNL Model	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
localxpa00:													
stw	-0.0948	-0.0405	-0.0194	-0.0599	-0.0784	-0.0519	-0.1302	-0.0027	-0.0069	-0.0096	-0.0247	-0.0287	-0.0534
	(0.0723)	(0.0902)	(0.0353)	(0.1236)	(0.0898)	(0.0540)	(0.1341)	(0.0572)	(0.0850)	(0.1420)	(0.0721)	(0.0648)	(0.1349)
adj	-0.0846	-0.0176	-0.0076	-0.0252	0.0538	0.0341	0.0878	-0.0082	-0.0118	-0.0200	-0.0059	-0.0063	-0.0122
	(0.0538)	(0.0888)	(0.0360)	(0.1245)	(0.0959)	(0.0673)	(0.1592)	(0.0466)	(0.0645)	(0.1108)	(0.0614)	(0.0595)	(0.1208)
no	0.1794	0.0582	0.0270	0.0852	0.0246	0.0178	0.0424	0.0109	0.0187	0.0296	0.0306	0.0350	0.0656
	(0.0549)	(0.0561)	(0.0198)	(0.0696)	(0.0581)	(0.0355)	(0.0922)	(0.0465)	(0.0673)	(0.1132)	(0.0549)	(0.0428)	(0.0938)
localxpa_ch9000:													
stw	0.0912	0.0264	0.0145	0.0409	0.0773	0.0518	0.1291	-0.0033	0.0007	-0.0026	0.0221	0.0274	0.0494
	(0.1091)	(0.1193)	(0.0487)	(0.16714)	(0.1160)	(0.0740)	(0.1825)	(0.0721)	(0.1090)	(0.1808)	(0.0908)	(0.0878)	(0.1770)
adj	0.1090	0.0491	0.0206	0.0696	-0.0426	-0.0267	-0.0693	0.0225	0.0321	0.0546	0.0174	0.0178	0.0352
	(0.0861)	(0.1212)	(0.0500)	(0.1694)	(0.1270)	(0.0845)	(0.2096)	(0.0645)	(0.0865)	(0.1489)	(0.0827)	(0.0797)	(0.1616)
no	-0.2002	-0.0755	-0.0350	-0.1105	-0.0347	-0.0252	-0.0598	-0.0192	-0.0329	-0.0521	-0.0395	-0.0451	-0.0846
	(0.0764)	(0.0673)	(0.0304)	(0.0886)	(0.0693)	(0.0461)	(0.1127)	(0.0566)	(0.0819)	(0.1365)	(0.0649)	(0.0580)	(0.1164)
ptaxrate00:													
stw	-0.0052	-0.0037	-0.0015	-0.0052	-0.0018	-0.0011	-0.0029	-0.0016	-0.0022	-0.0038	-0.0021	-0.0020	-0.0041
	(0.0059)	(0.0048)	(0.0024)	(0.0070)	(0.0045)	(0.0047)	(0.0034)	(0.0034)	(0.0052)	(0.0083)	(0.0039)	(0.0042)	(0.0078)
adj	0.0062	0.0037	0.0015	0.0053	0.0019	0.0012	0.0031	0.0015	0.0021	0.0036	0.0020	0.0019	0.0039
	(0.0047)	(0.0046)	(0.0020)	(0.0064)	(0.0031)	(0.0031)	(0.0024)	(0.0028)	(0.0036)	(0.0062)	(0.0035)	(0.0032)	(0.0065)
no	-0.0010	-0.0000	-0.0000	-0.0000	-0.0001	-0.0001	-0.0002	0.0001	0.0001	0.0002	0.0001	0.0001	0.0002
	(0.0045)	(0.0034)	(0.0016)	(0.0050)	(0.0075)	(0.0076)	(0.0058)	(0.0026)	(0.0045)	(0.0071)	(0.0029)	(0.0033)	(0.0062)
pctag00:													
stw	0.0067	0.0072	0.0034	0.0106	0.0072	0.0053	0.0126	0.0072	0.0125	0.0197	0.0063	0.0073	0.0136
	(0.0026)	(0.0032)	(0.0029)	(0.0052)	(0.0033)	(0.0047)	(0.0064)	(0.0030)	(0.0122)	(0.0128)	(0.0029)	(0.0071)	(0.0082)
adj	0.0035	0.0019	0.0008	0.0027	0.0040	0.0028	0.0068	0.0014	0.0022	0.0037	0.0017	0.0018	0.0036
	(0.0017)	(0.0017)	(0.0008)	(0.0024)	(0.0036)	(0.0032)	(0.0062)	(0.0019)	(0.0022)	(0.0038)	(0.0019)	(0.0016)	(0.0031)
no	-0.0102	-0.0091	-0.0042	-0.0133	-0.0112	-0.0081	-0.0193	-0.0086	-0.0148	-0.0234	-0.0080	-0.0092	-0.0172
	(0.0023)	(0.0034)	(0.0034)	(0.0054)	(0.0043)	(0.0040)	(0.0084)	(0.0037)	(0.0091)	(0.0120)	(0.0034)	(0.0046)	(0.0078)

Table 5-5. Continued

		SMNL Model (W_3nn)			SMNL Model (W_6nn)			SMNL Model (W_8nn)			SMNL Model (W_10nn)		
	MNL Model	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
pctag_ch9000:													
stw	-0.0071	-0.0032	-0.0010	-0.0043	0.0005	0.0009	0.0015	0.0010	0.0031	0.0041	-0.0012	-0.0004	-0.0016
	(0.0088)	(0.0076)	(0.0035)	(0.0110)	(0.0069)	(0.0050)	(0.0119)	(0.0063)	(0.0116)	(0.0178)	(0.0068)	(0.0074)	(0.0143)
adj	0.0110	0.0091	0.0037	0.0128	0.0089	0.0059	0.0148	0.0049	0.0071	0.0120	0.0066	0.0065	0.0131
	(0.0057)	(0.0048)	(0.0028)	(0.0062)	(0.0057)	(0.0053)	(0.0093)	(0.0051)	(0.0042)	(0.0078)	(0.0053)	(0.0039)	(0.0069)
no	-0.0039	-0.0058	-0.0027	-0.0085	-0.0094	-0.0068	-0.0162	-0.0060	-0.0102	-0.0161	-0.0054	-0.0062	-0.0116
	(0.0080)	(0.0080)	(0.0040)	(0.0115)	(0.0081)	(0.0076)	(0.0143)	(0.0066)	(0.0123)	(0.0171)	(0.0069)	(0.0084)	(0.0143)
cl2valpera00:													
stw	-0.0006	0.0026	0.0009	0.0036	-0.0000	-0.0001	-0.0001	0.0006	0.0007	0.0014	-0.0000	-0.0002	-0.0002
	(0.0024)	(0.0036)	(0.0016)	(0.0051)	(0.0033)	(0.0021)	(0.0054)	(0.0021)	(0.0030)	(0.0051)	(0.0026)	(0.0026)	(0.0052)
adj	-0.0031	-0.0051	-0.0021	-0.0071	-0.0013	-0.0009	-0.0022	-0.0014	-0.0019	-0.0033	-0.0012	-0.0012	-0.0024
	(0.0021)	(0.0040)	(0.0020)	(0.0054)	(0.0037)	(0.0024)	(0.0061)	(0.0022)	(0.0027)	(0.0047)	(0.0027)	(0.0026)	(0.0051)
no	0.0037	0.0025	0.0011	0.0036	0.0013	0.0009	0.0022	0.0007	0.0012	0.0020	0.0012	0.0014	0.0026
	(0.0015)	(0.0016)	(0.0010)	(0.0022)	(0.0015)	(0.0012)	(0.0026)	(0.0012)	(0.0019)	(0.0030)	(0.0014)	(0.0016)	(0.0026)

Note: Standard errors of the estimated marginal effects presented in parentheses are obtained using the Delta method.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

The main contribution of this dissertation is methodological. The linearization estimation methodology of Klier and McMillen (2008)—originally applied to a binary logit model—is extended to a multinomial setting for various specifications of logit models, all of which have different empirical applications. The performance of this methodology in finite samples is assessed in the context of a spatial multinomial logit (SMNL) model using Monte Carlo methods. Simulation results indicate that the linearized model provides a good approximation to the original spatial model and produces fairly accurate estimates for a reasonable range of induced spatial dependence in the simulated data. Judging from these results, the linearization approach appears successful. We further note that the finite sample properties of other spatial estimators proposed in this dissertation, namely the spatial conditional logit (SCL) and the mixed logit (MXL) estimators, were assessed as well, although the results were not presented here as they were qualitative similar to the results from the SMNL estimator. However, further examination of the performance of the spatial ordered logit (SOL) estimator is needed because of its distinctive ordered structure.

There are several advantages to the estimation methodology proposed in this dissertation relative to other spatial methods and standard modeling methods. First, unlike the majority of the spatial estimators in the literature, it is designed to model spatially dependent polychotomous choice decisions. Second, it allows a model with a spatially lagged dependent variable to be estimated within a discrete-choice framework even with very large samples, which is typical in micro-level data. Third, the Monte Carlo evidence for the SMNL estimator indicates that this estimator performs well in

capturing the spatial dependence in the simulated data. Finally, the spatial estimation method enables the estimation of indirect effects (in addition to direct effects), providing abundant information about spatial linkages and the strength of potential interaction between economic agents at different locations. This information is of value to inform policy.

The proposed spatial estimation approach is used in two empirical studies. The first study involves a spatial analysis of landowners' land-use conversion decisions. The main objective of this study is to understand what drives land-use conversion and identify the factors that play a significant role in the conversion of land to urban use. We estimate a spatially-explicit model of land-use conversion in the rural-urban fringe using parcel-level data from a rural-urban fringe county in Ohio, Medina County; a data scale that is appropriate for modeling the economic decision of the individual landowners. Unlike previous spatially-explicit studies, four land-use categories are considered in the analysis: agricultural, residential, industrial, and commercial. Empirical results from this application corroborate previous research findings suggesting that the location of new urban development is guided by a preference for lower density areas but in proximity to current urban development. This development trend is sometimes deterred by zoning policies. The main insight gained with this application is that spatial dependence is an important factor to take into account when analyzing land use conversion decisions because it helps us understand the underlying mechanisms of land-use changes. We find significant evidence of spatial dependence in land-use decisions; consistent with the notion that land-use change is a spatial process. The presence of spatial spillover effects suggests that local policies designed at a small scale could lead to sub-optimal

land-use patterns at a regional level. Thus, knowledge of the presence and the extent of spatial spillover effects can help inform the design of land use policy at an appropriate scale.

An important area to extend this analysis is to look at the predictions of future changes in land-use patterns. An initial assessment of the predictive ability of the model can be done by performing categorical in-sample predictions. Better in sample predictions (relative to the non-spatial model) are likely to translate to better predictions for future land-use changes. The results from the model can also be used with other spatial analysis tools such as ArcGIS to perform scenario analysis and examine how a change in one of the factors that influences land-use conversion changes the predicted land-use patterns.

The focus of the second study is neighborhood influence in the adoption of public policies. It uses school district data to estimate a spatially-explicit model of policy adoption decisions of Ohio's school districts. The policy under consideration is interdistrict open enrollment with the following policy alternatives; statewide open enrollment, adjacent open enrollment, and no open enrollment. The empirical results substantiate extant descriptive evidence regarding the determinants of adoption of open enrollment policies. Among the most influential factors are minority enrollment, racial and income heterogeneity, changes in student-teacher ratio, student academic performance, and financial factors. More importantly, the results show strong neighborhood influence in policy adoption decisions. The results from this analysis provide further support to the view that public policy adoption is stimulated by interaction among decision makers at a local level. In addition, the adoption of new public policies

much like the diffusion of technology innovations is largely influenced by reference groups. A spatial analysis of these neighborhood effects can help identify the most influential locations.

Other directions in which this analysis can be extended is to analyze the school districts' adoption decisions not only over space, but also over time. In addition, other ways in which neighbors may influence each other's adoption decisions can be explored since our specification of the spatial weights matrix assumes that the spatial influence is mutually symmetric. Another interesting venue of future work would be to combine the analyses from both empirical studies to examine the school-housing connection.

Finally, the methodology extended within this dissertation is relevant to myriad empirical problems that require micro-level analysis and involve a discrete-choice decision in which relative location matters. For instance, it can be employed to model household location decision, firm location decisions, and occupational choice (based on "economic distance" measures), among others.

APPENDIX A
LOGIT CHOICE PROBABILITIES AND MARGINAL EFFECTS

Derivation of Logit Probabilities

Consider the RUM equations:

$$U_{il} = V_{il} + \varepsilon_{il}, \qquad \forall l \in C, C = \{1, \dots, L\} \tag{A.1}$$

where U_{il} is individual i's indirect utility from alternative l; V_{il} is the systematic component of the indirect utility for individual i associated with alternative l; and ε_{il} is the unobserved component of utility for individual i associated with alternative l. In this set up, individual i chooses a particular alternative from the choice set, say k, if and only if:

$$U_{ik} > U_{il}, \forall k \neq l, l \in C \tag{A.2}$$

The probability that individual i chooses alternative k is given by:

$$P_{ik} = P(U_{ik} > U_{il}, \forall k \neq l) = P(V_{ik} + \varepsilon_{ik} > V_{il} + \varepsilon_{il}, \forall k \neq l)$$

$$= P(\varepsilon_{il} - \varepsilon_{ik} < V_{ik} - V_{il}, \forall k \neq l) = \int_{\varepsilon} I(\varepsilon_{il} - \varepsilon_{ik} < V_{ik} - V_{il}, \forall k \neq l) f(\varepsilon) d\varepsilon$$

where $f(\varepsilon) = f(\varepsilon_{i1}, \dots \varepsilon_{iL})$. Therefore,

$$P_{ik} = \int_{\varepsilon} I[\varepsilon_{il} < \varepsilon_{ik} + (V_{ik} - V_{il}), \forall k \neq l] f(\varepsilon) d\varepsilon \tag{A.3}$$

Let alternative k be the first alternative in the choice set and write the probability P_{i1} more explicitly as:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} \int_{\varepsilon_{i2=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i2})} \int_{\varepsilon_{i3=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i3})} \dots$$

$$\int_{\varepsilon_{i(l-1)=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})} \int_{\varepsilon_{il=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{il})} f(\varepsilon_{i1}, \dots \varepsilon_{il}) d\varepsilon_{il} d\varepsilon_{i(l-1)} \dots d\varepsilon_{i2} d\varepsilon_{i1}$$

Since the error terms are assumed to be independently distributed, then:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} f(\varepsilon_{i1}) \int_{\varepsilon_{i2=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i2})} f(\varepsilon_{i2}) d\varepsilon_{i2} \int_{\varepsilon_{i3=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i3})} f(\varepsilon_{i3}) d\varepsilon_{i3} \ldots$$

$$\int_{\varepsilon_{i(l-1)=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})} f(\varepsilon_{i(l-1)}) d\varepsilon_{i(l-1)} \int_{\varepsilon_{il=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{il})} f(\varepsilon_{il}) d\varepsilon_{il} d\varepsilon_{i1}$$

where $f(\varepsilon_l)$ is the PDF of the error term for alternative l. Moreover, since the error terms are also assumed to be distributed type I extreme value (or Gumbel), the PDF takes the following functional form: $f(\varepsilon_{il}) = e^{-\varepsilon_{il}} e^{-e^{-\varepsilon_{il}}}$. Thus,

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{-\varepsilon_{i1}} e^{-e^{-\varepsilon_{i1}}} \int_{\varepsilon_{i2=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i2})} e^{-\varepsilon_{i2}} e^{-e^{-\varepsilon_{i2}}} d\varepsilon_{i2} \int_{\varepsilon_{i3=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i3})} e^{-\varepsilon_{i3}} e^{-e^{-\varepsilon_{i3}}} d\varepsilon_{i3} \ldots$$

$$\int_{\varepsilon_{i(l-1)=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})} e^{-\varepsilon_{i(l-1)}} e^{-e^{-\varepsilon_{i(l-1)}}} d\varepsilon_{i(l-1)} \int_{\varepsilon_{il=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{il})} e^{-\varepsilon_{il}} e^{-e^{-\varepsilon_{il}}} d\varepsilon_{il} d\varepsilon_{i1}$$

Note that $d(-e^{-\varepsilon_{il}}) = (-e^{-\varepsilon_{il}})(-1) d\varepsilon_{il} = e^{-\varepsilon_{il}} d\varepsilon_{il}$. Hence, integrating by substitution:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{-e^{-\varepsilon_{i1}}} \int_{\varepsilon_{i2=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i2})} e^{-e^{-\varepsilon_{i2}}} d(-e^{-\varepsilon_{i2}}) \int_{\varepsilon_{i3=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i3})} e^{-e^{-\varepsilon_{i3}}} d(-e^{-\varepsilon_{i3}}) \ldots$$

$$\int_{\varepsilon_{i(l-1)=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})} e^{-e^{-\varepsilon_{i(l-1)}}} d(-e^{-\varepsilon_{i(l-1)}}) \int_{\varepsilon_{il=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{il})} e^{-e^{-\varepsilon_{il}}} d(-e^{-\varepsilon_{il}}) d(-e^{-\varepsilon_{i1}})$$

Note: $\int_a^b e^{-e^{-\varepsilon_{il}}} d(-e^{-\varepsilon_{il}}) = e^{-e^{-\varepsilon_{il}}} \,|_a^b$. Moreover, as $\varepsilon_{il} \to -\infty, e^{-\varepsilon_{il}} \to \infty,$ and $e^{-e^{-\varepsilon_{il}}} \to 0$.

Thus, evaluating the definite integrals:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{-e^{-\varepsilon_{i1}}} \left[[e^{-e^{-\varepsilon_{i2}}}]_{\varepsilon_{i2=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i2})} \right] \left[[e^{-e^{-\varepsilon_{i3}}}]_{\varepsilon_{i3=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i3})} \right] \ldots$$

$$\left[[e^{-e^{-\varepsilon_{i(l-1)}}}]_{\varepsilon_{i(l-1)=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})} \right] \left[[e^{-e^{-\varepsilon_{il}}}]_{\varepsilon_{il=-\infty}}^{\varepsilon_{i1}+(V_{i1}-V_{il})} \right] d(-e^{-\varepsilon_{i1}})$$

Evaluating the integrals and simplifying:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{-e^{-\varepsilon_{i1}}} e^{-e^{-[\varepsilon_{i1}+(V_{i1}-V_{i2})]}} e^{-e^{-[\varepsilon_{i1}+(V_{i1}-V_{i3})]}} \ldots$$

$$e^{-e^{-\left[\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})\right]}} e^{-e^{-[\varepsilon_{i1}+(V_{i1}-V_{il})]}} d(-e^{-\varepsilon_{i1}})$$

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{-e^{-\varepsilon_{i1}} - e^{-[\varepsilon_{i1}+(V_{i1}-V_{i2})]} - e^{-[\varepsilon_{i1}+(V_{i1}-V_{i3})]} \cdots - e^{-\left[\varepsilon_{i1}+(V_{i1}-V_{i(l-1)})\right]} - e^{-[\varepsilon_{i1}+(V_{i1}-V_{il})]}} \, d(-e^{-\varepsilon_{i1}})$$
$$= \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{(-e^{-\varepsilon_{i1}})\left(1+e^{-(V_{i1}-V_{i2})}+e^{-(V_{i1}-V_{i3})\cdots}+e^{-(V_{i1}-V_{i(l-1)})}+e^{-(V_{i1}-V_{il})}\right)} \, d(-e^{-\varepsilon_{i1}})$$

To simplify the notation, let $a = 1 + e^{-(V_{i1}-V_{i2})} + e^{-(V_{i1}-V_{i3})\cdots} + e^{-(V_{i1}-V_{i(l-1)})} + e^{-(V_{i1}-V_{il})}$.

P_{i1} can be written more compactly as:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} e^{(-e^{-\varepsilon_{i1}})a} \, d(-e^{-\varepsilon_{i1}})$$

P_{i1} can be integrated by substitution by noting that $d((-e^{-\varepsilon_{i1}})a) = ad(-e^{-\varepsilon_{i1}})$, thus $d(-e^{-\varepsilon_{i1}}) = \frac{1}{a}\, d((-e^{-\varepsilon_{i1}})a)$. Substituting it in the probability formula:

$$P_{i1} = \int_{\varepsilon_{i1=-\infty}}^{\infty} \frac{1}{a} e^{(-e^{-\varepsilon_{i1}})a} \, d\left((-e^{-\varepsilon_{i1}})a\right) = \left.\frac{e^{(-e^{-\varepsilon_{i1}})a}}{a}\right|_{-\infty}^{\infty} = \frac{1}{a}$$

because as $\varepsilon_{i1} \to -\infty, e^{-\varepsilon_{i1}} \to \infty,\ e^{-e^{-\varepsilon_{i1}}} \to 0$ and as $\varepsilon_{i1} \to \infty, e^{-\varepsilon_{i1}} \to 0,\ e^{-e^{-\varepsilon_{i1}}} \to 1$.

Substituting back the expression for a:

$$P_{i1} = \frac{1}{1 + e^{-(V_{i1}-V_{i2})} + e^{-(V_{i1}-V_{i3})\cdots} + e^{-(V_{i1}-V_{i(l-1)})} + e^{-(V_{i1}-V_{il})}}$$
$$= \frac{1}{1 + e^{-V_{i1}}(e^{V_{i2}} + e^{V_{i3}} + \cdots + e^{V_{i(l-1)}} + e^{V_{il}})}$$
$$= \frac{e^{V_{i1}}}{(e^{V_{i1}} + e^{V_{i2}} + e^{V_{i3}} + \cdots + e^{V_{i(l-1)}} + e^{V_{il}})}$$
$$= \frac{e^{V_{i1}}}{\sum_l e^{V_{il}}}$$

Replacing alternative 1 with the general alternative k, we get the logit probability formula:

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_l e^{V_{il}}} \tag{A.4}$$

Marginal Effects and Elasticities

Marginal effects measure the change in the choice probability for a ceteris paribus change in an observed factor[1]. Let S_{il} denote an alternative-varying regressor that enters the systematic component of the utility (V_{il}) in (A.1). From (A.4), the probability that individual i chooses a particular alternative k is given by:

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_l e^{V_{il}}}$$

Hence, the change in the probability that decision maker i chooses alternative k given a change in an attribute associated with alternative k is given by the following marginal effect:

$$ME_{ikk} = \frac{\partial P_{ik}}{\partial S_{ik}} = \frac{\partial}{\partial S_{ik}}\left(\frac{e^{V_{ik}}}{\sum_l e^{V_{il}}}\right) = \frac{e^{V_{ik}}}{\sum_l e^{V_{il}}}\frac{\partial V_{ik}}{\partial S_{ik}} - \frac{e^{V_{ik}}}{(\sum_l e^{V_{il}})^2} e^{V_{ik}} \frac{\partial V_{ik}}{\partial S_{ik}} = \frac{\partial V_{ik}}{\partial S_{ik}}(P_{ik} - P_{ik}^2)$$

$$ME_{ikk} = \frac{\partial V_{ik}}{\partial S_{ik}} P_{ik}(1 - P_{ik}) \tag{A.5}$$

If the observed factors enter linearly in the systematic component of the utility, then $\frac{\partial V_{ik}}{\partial S_{ik}}$ is constant and equal to the model parameter associated with S_{ik}. Note that this marginal effect is the largest when $P_{ik} = 0.5$, so the effect of an attribute change is largest when there is a high degree of uncertainty regarding the choice. This observation is important for policy purposes, because it identifies the choices that could be mostly influenced by policy[2].

[1] This section derives marginal effects associated with changes in alternative-varying regressors. The computation of marginal effects for alternative-invariant regressors, however, is straightforward and results in similar expressions.

[2] See Train (2007) for policy examples.

Sometimes it is of interest to look at the change in the probability of choosing alternative k as an attribute of another alternative, say m, changes. In this case, the (cross) marginal effect is given by:

$$ME_{ikm} = \frac{\partial P_{ik}}{\partial S_{im}} = \frac{\partial}{\partial S_{im}}\left(\frac{e^{V_{ik}}}{\sum_l e^{V_{il}}}\right) = \frac{e^{V_{ik}}}{\sum_l e^{V_{il}}}\frac{\partial V_{ik}}{\partial S_{im}} - \frac{e^{V_{ik}}}{(\sum_l e^{V_{il}})^2} e^{V_{im}} \frac{\partial V_{im}}{\partial S_{im}}$$

$$ME_{ikm} = -\frac{\partial V_{im}}{\partial S_{im}} P_{ik} P_{im} \tag{A.6}$$

An interesting aspect of the choice probabilities is that, when an attribute of an alternative changes, the changes in the choice probabilities for all alternatives sum to zero (because the choice probabilities sum to one). Differently put, if the choice probability of an alternative with an improved attribute increases, it does so at the expense of the other alternatives. Generalizing the marginal effect expressions in (A.5) and (A.6) for any alternative l in the choice set and summing the marginal effects, we get:

$$\sum_{l=1}^{L} \frac{\partial P_{il}}{\partial S_{ik}} = \frac{\partial V_{ik}}{\partial S_{ik}} P_{ik}(1 - P_{ik}) + \sum\nolimits_{k \neq l} \left(-\frac{\partial V_{ik}}{\partial S_{ik}}\right) P_{ik} P_{il}$$

$$= \frac{\partial V_{ik}}{\partial S_{ik}} P_{ik} \left[(1 - P_{ik}) - \sum_{k \neq l} P_{il}\right]$$

$$= \frac{\partial V_{ik}}{\partial S_{ik}} P_{ik} [(1 - P_{ik}) - (1 - P_{ik})] = 0$$

$$\sum_{l=1}^{L} \frac{\partial P_{il}}{\partial S_{ik}} = 0 \tag{A.7}$$

Elasticities measure the percentage change in the choice probabilities for a percentage change in the value of a regressor. The (own) elasticity of a change in the

probability that decision maker i chooses alternative k given a change in an attribute associated with alternative k is:

$$E_{ikk} = \frac{\partial P_{ik}}{\partial S_{ik}} \frac{S_{ik}}{P_{ik}} = \frac{\partial V_{ik}}{\partial S_{ik}} P_{ik}(1 - P_{ik}) \frac{S_{ik}}{P_{ik}} = \frac{\partial V_{ik}}{\partial S_{ik}} S_{ik}(1 - P_{ik}) \tag{A.8}$$

If the observed factors linearly enter in the systematic component of the utility, then $\frac{\partial V_{ik}}{\partial S_{ik}}$ in (A.8) is also constant and equal to the model parameter associated with S_{ik}. Similarly, the (cross) elasticity of a change in the choice probability of alternative k as an attribute of alternative m changes, is:

$$E_{ikm} = \frac{\partial P_{ik}}{\partial S_{im}} \frac{S_{im}}{P_{ik}} = -\frac{\partial V_{im}}{\partial S_{im}} P_{ik} P_{im} \frac{S_{im}}{P_{ik}} = -\frac{\partial V_{im}}{\partial S_{im}} S_{im} P_{im} \tag{A.9}$$

Note that the cross-elasticity in (A.9) depends only on alternative m. Thus, a change in an attribute for alternative m changes the probabilities for all other alternatives by the same percentage (IIA property).

APPENDIX B
DERIVATION OF GRADIENTS FOR SPATIAL LOGIT MODELS

This appendix contains the derivations of the gradient terms for a spatial autoregressive lag (SAL) specification of the multinomial logit (MNL), conditional logit (CL), mixed logit (MXL), and ordered logit (OL) models[1]. The linearization methodology proposed in Chapter 3 for the estimation of these spatial discrete-choice models is plausible only if the gradient of the spatial autoregressive parameter(ρ) is nonzero once we linearize the model around a convenient point of initial parameter values which sets the spatial parameter equal to zero. Thus, in addition to obtaining the functional form for each of the gradients, we want to show that, at the linearization point, the spatial parameter remains identified from the corresponding gradient. For the spatial logit models with unordered choice alternatives, we start the gradient derivations with the mixed model since it combines both types of covariates (individual-specific and alternative-specific) and derive the spatial MNL and spatial CL models as special cases.

Spatial Mixed Logit (SMXL) Model

Consider the following SAL specification of a discrete-choice model:

$$Y_{ik}^{*} = \rho \sum_{j=1}^{N} w_{ij} Y_{jk}^{*} + X_i \beta_k + S_{ik}\gamma + \varepsilon_{ik} \tag{B.1}$$

with,

$$Y_i = k \quad \text{if} \quad Y_{ik}^{*} > Y_{il}^{*}, \forall k \neq l, l = \{1, \dots, L\}$$

where ${Y_k}^{*}$ is a latent dependent variable with observable counterpart Y; w_{ij} denotes the spatial weights relating observations i and j; X and S_k denote alternative-invariant and

[1] The difference between the polychotomous models with unordered choice alternatives is as follows. The MNL model accommodates only explanatory variables that vary over observations (individuals) but not over alternatives. The CL model accommodates only alternative-varying explanatory variables. The MXL combines the two preceding models. The terminology for the mixed logit model is not to be confused with the random parameters logit model (RPL) also often referred to as a mixed logit model.

alternative-varying covariates, respectively; ε_k is a vector of IID disturbances for alternative k; ρ is the spatial autoregressive parameter; and β_k and γ are the model parameters. The alternative with the highest latent utility is the one chosen by the decision maker, which is the choice observed in Y_i.

The model can be written in a reduced form as:

$$Y_{ik}^* = \sum_{j=1}^{N} \psi_{ij}\, X_j \beta_k + \sum_{j=1}^{N} \psi_{ij}\, S_{jk}\gamma + \sum_{j=1}^{N} \psi_{ij}\, \varepsilon_{jk} \tag{B.2}$$

with,

$$Y_i = k \quad \text{if} \quad Y_{ik}^* > Y_{il}^*, \forall k \neq l$$

where ψ_{ij} are the (i,j) elements of the spatial matrix $(I - \rho W)^{-1}$. The reduced model can be written in matrix notation as:

$$Y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} S\gamma + e, \qquad e = (I - \rho W)^{-1}\varepsilon \tag{B.3}$$

which results in the error covariance matrix: $V(e) \propto [(I - \rho W)'(I - \rho W)]^{-1}$. This covariance structure implies that the error terms (e) are both autocorrelated and heteroskedastic. Denote by σ_i^2 the variance of the error terms given by the diagonal elements of $V(e)$. The model can be normalized for heteroskedastic errors as follows:

$$\frac{Y_{ik}^*}{\sigma_i} = \frac{\sum_j \psi_{ij}\, X_j \beta_k}{\sigma_i} + \frac{\sum_j \psi_{ij}\, S_{jk}\gamma}{\sigma_i} + \frac{\sum_j \psi_{ij}\, \varepsilon_{jk}}{\sigma_i} \tag{B.4}$$

with,

$$Y_i = k \quad \text{if} \quad \frac{Y_{ik}^*}{\sigma_i} > \frac{Y_{il}^*}{\sigma_i}, \forall k \neq l$$

Define the dependent variable Y_i in a binary form as: $d_{ik} = I(Y_i = k)$. The model in (B.4) implies the following probability of choosing alternative k (by individual i):

$$P_{ik} = P(Y_i = k | X_i, S_{ik}) = P(d_{ik} = 1 | X_i, S_{ik}) = P\left(\frac{Y_{ik}^*}{\sigma_i} > \frac{Y_{il}^*}{\sigma_i}\right)$$

$$= P[(X_i^{**}\beta_k + S_{ik}^{**}\gamma + \varepsilon_{ik}^{**}) > (X_i^{**}\beta_l + S_{il}^{**}\gamma + \varepsilon_{il}^{**})]$$

Hence,

$$P_{ik} = P[\varepsilon_{il}^{**} - \varepsilon_{ik}^{**} < (X_i^{**}\beta_k + S_{ik}^{**}\gamma) - (X_i^{**}\beta_l + S_{il}^{**}\gamma)] \tag{B.5}$$

where $X^{**} = (I - \rho W)^{-1}X^*$, $X_i^* = X_i/\sigma_i$; $S^{**} = (I - \rho W)^{-1}S^*$, $S_{ik}^* = S_{ik}/\sigma_i$, $S_{il}^* = S_{il}/\sigma_i$; and $\varepsilon^{**} = (I - \rho W)^{-1}\varepsilon^*$, $\varepsilon_{ik}^* = \varepsilon_{ik}/\sigma_i$. If the error terms are assumed to be distributed IID type I extreme value, the error differences will possess a logistic distribution. These distributional assumptions, give rise to the SMXL model and the choice probabilities take the following functional form:

$$P_{ik} = \frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}} \tag{B.6}$$

Having specified the choice probabilities, the model gradients can be derived by differentiating these probabilities with respect to the model parameters(β_k, γ, ρ).

- Gradient of β_k:

$$\frac{\partial P_{ik}}{\partial \beta_k} = \frac{\left[e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}X_i^{**}\right]\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2} - \frac{\left(e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}\right)\left[e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}X_i^{**}\right]}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2}$$

$$= \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)X_i^{**} - \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)^2 X_i^{**}$$

$$= P_{ik}X_i^{**} - P_{ik}^2X_i^{**}$$

$$= P_{ik}(1 - P_{ik})X_i^{**}$$

For a general alternative $l \neq k$:

$$\frac{\partial P_{il}}{\partial \beta_k} = -\frac{\left(e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)\left[e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}X_i^{**}\right]}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2}$$

$$= -\left(\frac{e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)\left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)X_i^{**}$$

$$= -P_{il}P_{ik}X_i^{**}$$

Let $\delta_{ilk} = 1$ if $l = k$, and zero otherwise. Combining the results, the gradient term for β_k is given by:

$$G_{\beta_{ik}} = P_{ik}(\delta_{ilk} - P_{il})X_i^{**},\ \delta_{ilk} = I(l = k) \tag{B.7}$$

- Gradient of γ:

$$\frac{\partial P_{ik}}{\partial \gamma} = \frac{\left[(e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)})S_{ik}^{**}\right]\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2} - \frac{\left(e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}\right)\left[\sum_l \left(e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}S_{il}^{**}\right)\right]}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2}$$

$$= \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)S_{ik}^{**} - \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)\left[\frac{\sum_l (e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}S_{il}^{**})}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right]$$

$$= P_{ik}S_{ik}^{**} - P_{ik}\sum_l P_{il}\,S_{il}^{**}$$

$$= P_{ik}\left[S_{ik}^{**} - \sum_l P_{il}\,S_{il}^{**}\right]$$

Hence,

$$G_{\gamma i} = P_{ik}\left[S_{ik}^{**} - \sum_l P_{il}S_{il}^{**}\right] \tag{B.8}$$

- Gradient of ρ:

$$\frac{\partial P_{ik}}{\partial \rho} = \frac{\left[e^{\left(X_i^{**}\beta_k + S_{ik}^{**}\gamma\right)}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k + \left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right)\right]\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2}$$

$$- \frac{\left(e^{\left(X_i^{**}\beta_k + S_{ik}^{**}\gamma\right)}\right)\left[\sum_l\left(e^{\left(X_i^{**}\beta_l + S_{il}^{**}\gamma\right)}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l + \left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right)\right]}{\left(\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}\right)^2}$$

$$= \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k + \left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right)$$

$$- \left(\frac{e^{(X_i^{**}\beta_k + S_{ik}^{**}\gamma)}}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right)\left[\frac{\sum_l\left(e^{\left(X_i^{**}\beta_l + S_{il}^{**}\gamma\right)}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l + \left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right)}{\sum_l e^{(X_i^{**}\beta_l + S_{il}^{**}\gamma)}}\right]$$

$$= P_{ik}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k + \left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right) - P_{ik}\left[\sum_l P_{il}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l + \left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right]$$

$$= P_{ik}\left[\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k + \left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right) - \left[\sum_l P_{il}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l + \left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right]\right]$$

In order to obtain the partial derivatives $\frac{\partial X_i^{**}}{\partial \rho}$, $\frac{\partial S_{ik}^{**}}{\partial \rho}$, and $\frac{\partial S_{il}^{**}}{\partial \rho}$, recall that

$X^{**} = (I - \rho W)^{-1}X^*$ and $S^{**} = (I - \rho W)^{-1}S^*$.

$$\frac{\partial X^{**}}{\partial \rho} = \frac{\partial[(I - \rho W)^{-1}]}{\partial \rho}X^* + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho}$$

$$= -(I - \rho W)^{-1}\left[\frac{\partial(I - \rho W)}{\partial \rho}\right](I - \rho W)^{-1}X^* + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho}$$

$$= -(I - \rho W)^{-1}(-W)(I - \rho W)^{-1}X^* + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho}$$

$$= (I - \rho W)^{-1}W(I - \rho W)^{-1}X^* + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho}$$

$$= H + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho}$$

Hence,

$$\frac{\partial X^{**}}{\partial \rho} = H + (I - \rho W)^{-1}\frac{\partial X^*}{\partial \rho} \tag{B.9}$$

where $H = (I - \rho W)^{-1}W(I - \rho W)^{-1}X^*$.

Similarly,

$$\frac{\partial S^{**}}{\partial \rho} = \frac{\partial[(I - \rho W)^{-1}]}{\partial \rho}S^* + (I - \rho W)^{-1}\frac{\partial S^*}{\partial \rho}$$

$$= (I - \rho W)^{-1}W(I - \rho W)^{-1}S^* + (I - \rho W)^{-1}\frac{\partial S^*}{\partial \rho}$$

$$= K + (I - \rho W)^{-1}\frac{\partial S^*}{\partial \rho}$$

$$\frac{\partial S^{**}}{\partial \rho} = K + (I - \rho W)^{-1} \frac{\partial S^{*}}{\partial \rho} \tag{B.10}$$

where $K = (I - \rho W)^{-1} W (I - \rho W)^{-1} S^{*}$.

To complete the derivation, we need to compute the partial derivatives $\frac{\partial X^{*}}{\partial \rho}$ and $\frac{\partial S^{*}}{\partial \rho}$ in (B.9) and (B.10). Note that $X_i^{*} = X_i/\sigma_i$, $S_{ik}^{*} = S_{ik}/\sigma_i$, and $S_{il}^{*} = S_{il}/\sigma_i$, which can be written in terms of the error variances as follows: $X_i^{*} = X_i/\sqrt{\sigma_i^2}$, $S_{ik}^{*} = S_{ik}/\sqrt{\sigma_i^2}$, and $S_{il}^{*} = S_{il}/\sqrt{\sigma_i^2}$. Hence,

$$\frac{\partial X_i^{*}}{\partial \rho} = \frac{\partial X_i^{*}}{\partial \sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho} = -\frac{1}{2}\frac{X_i}{(\sigma_i^2)^{3/2}}\frac{\partial \sigma_i^2}{\partial \rho} = -\frac{1}{2}\frac{X_i^{*}}{\sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho}$$

Similarly,

$$\frac{\partial S_{ik}^{*}}{\partial \rho} = \frac{\partial S_{ik}^{*}}{\partial \sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho} = -\frac{1}{2}\frac{S_{ik}^{*}}{\sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho}$$

$$\frac{\partial S_{il}^{*}}{\partial \rho} = \frac{\partial S_{il}^{*}}{\partial \sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho} = -\frac{1}{2}\frac{S_{il}^{*}}{\sigma_i^2}\frac{\partial \sigma_i^2}{\partial \rho}$$

The last partial derivative, $\frac{\partial \sigma_i^2}{\partial \rho}$, can be obtained by differentiating the covariance matrix $V(e) = (I - \rho W)^{-1}(I - \rho W')^{-1}$ with respect to ρ and extracting the diagonal terms. Thus,

$$\begin{aligned}
\frac{\partial V(e)}{\partial \rho} &= \frac{\partial[(I - \rho W)^{-1}]}{\partial \rho}(I - \rho W')^{-1} + (I - \rho W)^{-1}\frac{\partial[(I - \rho W')^{-1}]}{\partial \rho} \\
&= -(I - \rho W)^{-1}(-W)(I - \rho W)^{-1}(I - \rho W')^{-1} \\
&\qquad - (I - \rho W)^{-1}(I - \rho W')^{-1}(-W')(I - \rho W')^{-1} \\
&= (I - \rho W)^{-1} W (I - \rho W)^{-1}(I - \rho W')^{-1} \\
&\qquad + (I - \rho W)^{-1}(I - \rho W')^{-1} W' (I - \rho W')^{-1}
\end{aligned}$$

$$= 2(I - \rho W)^{-1} W (I - \rho W)^{-1} (I - \rho W')^{-1}$$

$$= 2(I - \rho W)^{-1} W (I - \rho W)^{-1} (I - \rho W)^{-1}$$

because $(I - \rho W)^{-1} W (I - \rho W)^{-1} (I - \rho W')^{-1}$ and W are symmetric, hence equal to their transpose. Thus,

$$\frac{\partial \sigma_i^2}{\partial \rho} = 2\Lambda_{ii} \tag{B.11}$$

where $\Lambda = (I - \rho W)^{-1} W (I - \rho W)^{-1} (I - \rho W)^{-1}$.

Consequently,

$$\begin{aligned} \frac{\partial X_i^*}{\partial \rho} &= -\frac{X_i^*}{\sigma_i^2} \Lambda_{ii} \\ \frac{\partial S_{ik}^*}{\partial \rho} &= -\frac{S_{ik}^*}{\sigma_i^2} \Lambda_{ii} \\ \frac{\partial S_{il}^*}{\partial \rho} &= -\frac{S_{il}^*}{\sigma_i^2} \Lambda_{ii} \end{aligned} \tag{B.12}$$

Combining the results in (B.9), (B.10), and (B.12):

$$\begin{aligned} \frac{\partial X_i^{**}}{\partial \rho} &= H_i - \frac{X_i^{**}}{\sigma_i^2} \Lambda_{ii} \\ \frac{\partial S_{ik}^{**}}{\partial \rho} &= K_{ik} - \frac{S_{ik}^{**}}{\sigma_i^2} \Lambda_{ii} \\ \frac{\partial S_{il}^{**}}{\partial \rho} &= K_{il} - \frac{S_{il}^{**}}{\sigma_i^2} \Lambda_{ii} \end{aligned} \tag{B.13}$$

Substituting (B.13) into the gradient expression, the gradient of ρ becomes:

$$\begin{aligned} G_{\rho i} = P_{ik} \Bigg[\Bigg[\left(H_i \beta_k - \frac{X_i^{**} \beta_k}{\sigma_i^2} \Lambda_{ii} \right) + \left(K_{ik} \gamma - \frac{S_{ik}^{**} \gamma}{\sigma_i^2} \Lambda_{ii} \right) \Bigg] \\ - \sum_l P_{il} \Bigg[\left(H_i \beta_l - \frac{X_i^{**} \beta_l}{\sigma_i^2} \Lambda_{ii} \right) + \left(K_{il} \gamma - \frac{S_{il}^{**} \gamma}{\sigma_i^2} \Lambda_{ii} \right) \Bigg] \Bigg] \end{aligned} \tag{B.14}$$

where $H = (I - \rho W)^{-1} W (I - \rho W)^{-1} X^*$, $K = (I - \rho W)^{-1} W (I - \rho W)^{-1} S^*$, and $\Lambda = (I - \rho W)^{-1} W (I - \rho W)^{-1} (I - \rho W)^{-1}$.

At the linearization point ($\rho = 0$), $X_i^{**} = X_i^* = X_i$, $S_{ik}^{**} = S_{ik}^* = S_{ik}$, $S_{il}^{**} = S_{il}^* = S_{il}$, and $\Lambda_{ii} = w_{ii}$, where $w_{ii} = 0$ by construction, thus the terms containing Λ_{ii} in (B.14) vanish. However, the gradient of ρ does not vanish because of the terms H and K which become WX and WS, respectively. Thus, this model can be linearized around a point of parameter values which sets $\rho = 0$ because the parameter ρ remains identified from the corresponding gradient. At the linearization point, the gradients of the linearized SMXL model simplify to:

$$
\begin{aligned}
&G_{\beta_{ik}} = P_{ik}(\delta_{ilk} - P_{il})X_i,\ \delta_{ilk} = I(l = k) \\
&G_{\gamma i} = P_{ik}\left[S_{ik} - \sum_l P_{il}S_{il}\right] \\
&G_{\rho i} = P_{ik}\left[\left[(WX)_i\beta_k + (WS)_{ik}\gamma\right] - \sum_l P_{il}\left[(WX)_i\beta_l + (WS)_{il}\gamma\right]\right]
\end{aligned}
\tag{B.15}
$$

Spatial Multinomial Logit (SMNL) Model

A SAL specification for the MNL model can be obtained by setting $\gamma = 0$ in (B.1) as follows:

$$
Y_{ik}^* = \rho \sum_{j=i}^{N} w_{ij}Y_{jk}^* + X_i\beta_k + \varepsilon_{ik}
\tag{B.16}
$$

with,

$$
Y_i = k \quad \text{if} \quad Y_{ik}^* > Y_{il}^*, \forall k \neq l, l = \{1, \dots, L\}
$$

where Y_k^* is a latent dependent variable with observable counterpart Y; X denotes covariates that are alternative-invariant; w_{ij} denotes the spatial weights for observations i and j; ρ and β_k are the model parameters; and ε_k is a vector of IID disturbances for alternative k. Then, from (B.5) the probability that individual i chooses alternative k is given by:

$$P_{ik} = P(d_{ik} = 1|X_i) = P(\varepsilon_{il}^{**} - \varepsilon_{ik}^{**} < X_i^{**}\beta_k - X_i^{**}\beta_l\ ) \tag{B.17}$$

Assuming that ε_l^{**} and ε_k^{**} are distributed IID type I extreme value, the choice probabilities for the SMNL model take the functional form:

$$P_{ik} = \frac{e^{\,X_i^{**}\beta_k}}{\sum_l e^{\,X_i^{**}\beta_l}} \tag{B.18}$$

The gradients with respect to the model parameters (β_k,ρ) for this model are obtained as follows.

- Gradient of β_k:

$$\frac{\partial P_{ik}}{\partial \beta_k} = \frac{\left[(e^{\,X_i^{**}\beta_k})X_i^{**}\right]\left(\sum_l e^{\,X_i^{**}\beta_l}\right)}{\left(\sum_l e^{\,X_i^{**}\beta_l}\right)^2} - \frac{(e^{\,X_i^{**}\beta_k})\left[(e^{\,X_i^{**}\beta_k})X_i^{**}\right]}{\left(\sum_l e^{\,X_i^{**}\beta_l}\right)^2}$$

$$= \left(\frac{e^{\,X_i^{**}\beta_k}}{\sum_l e^{\,X_i^{**}\beta_l}}\right)X_i^{**} - \left(\frac{e^{\,X_i^{**}\beta_k}}{\sum_l e^{\,X_i^{**}\beta_l}}\right)^2 X_i^{**}$$

$$= P_{ik}(1 - P_{ik})X_i^{**}$$

For alternative $l \neq k$,

$$\frac{\partial P_{il}}{\partial \beta_k} = -\frac{\left(e^{\,X_i^{**}\beta_l}\right)\left[\left(e^{\,X_i^{**}\beta_k}\right)X_i^{**}\right]}{\left(\sum_l e^{\,X_i^{**}\beta_l}\right)^2}$$

$$= -\left(\frac{e^{\,X_i^{**}\beta_l}}{\sum_l e^{\,X_i^{**}\beta_l}}\right)\left(\frac{e^{\,X_i^{**}\beta_k}}{\sum_l e^{\,X_i^{**}\beta_l}}\right)X_i^{**}$$

$$= -P_{il}P_{ik}X_i^{**}$$

Define $\delta_{ilk} = 1$ if $l = k$, and zero otherwise. Then, the gradient term for β_k is given by:

$$G_{\beta_{ik}} = P_{ik}(\delta_{ilk} - P_{il})X_i^{**}, \delta_{ilk} = I(l = k) \tag{B.19}$$

- Gradient of ρ:

$$\frac{\partial P_{ik}}{\partial \rho} = \frac{\left[(e^{\,X_i^{**}\beta_k})\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k\right)\right]\left(\sum_l e^{\,X_i^{**}\beta_l}\right)}{\left(\sum_l e^{\,X_i^{**}\beta_l}\right)^2} - \frac{\left(e^{\,X_i^{**}\beta_k}\right)\left[\sum_l\left(e^{\,X_i^{**}\beta_l}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l\right)\right)\right]}{\left(\sum_l e^{\,X_i^{**}\beta_l}\right)^2}$$

$$= \left(\frac{e^{X_i^{**}\beta_k}}{\sum_l e^{X_i^{**}\beta_l}}\right)\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k\right) - \left(\frac{e^{X_i^{**}\beta_k}}{\sum_l e^{X_i^{**}\beta_l}}\right)\sum_l \left(\frac{e^{X_i^{**}\beta_l}}{\sum_l e^{X_i^{**}\beta_l}}\right)\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l\right)$$

$$= P_{ik}\left[\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_k\right) - \sum_l P_{il}\left(\left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta_l\right)\right]$$

From (B.13): $\frac{\partial X_i^{**}}{\partial \rho} = H_i - \frac{X_i^{**}}{\sigma_i^2}\Lambda_{ii}$. Hence, the gradient of ρ is:

$$G_{\rho_i} = P_{ik}\left[\left(H_i\beta_k - \frac{X_i^{**}\beta_k}{\sigma_i^2}\Lambda_{ii}\right) - \sum_l P_{il}\left(H_i\beta_l - \frac{X_i^{**}\beta_l}{\sigma_i^2}\Lambda_{ii}\right)\right] \tag{B.20}$$

where $H = (I - \rho W)^{-1}W(I - \rho W)^{-1}X^*$ and $\Lambda = (I - \rho W)^{-1}W(I - \rho W)^{-1}(I - \rho W)^{-1}$.

At the linearization point ($\rho = 0$), $X_i^{**} = X_i^* = X_i$ and $\Lambda_{ii} = w_{ii} = 0$, thus the terms containing Λ_{ii} in (B.20) vanish. The gradient of ρ - and thus, the spatial parameter ρ - remain identified because of the term H, which reduces to WX. Thus, this model can also be linearized around a point of parameter values which sets $\rho = 0$. The gradients for the linearized SMNL model are:

$$G_{\beta_{ik}} = P_{ik}(\delta_{ilk} - P_{il})X_i,\ \delta_{ilk} = I(l = k)$$

$$G_{\rho_i} = P_{ik}\left[(WX)_i\beta_k - \sum_l P_{il}(WX)_i\beta_l\right] \tag{B.21}$$

Spatial Conditional Logit (SCL) Model

A SAL specification for the CL model can be obtained by setting $\beta_k = 0$ in (B.1) as follows:

$$Y_{ik}^* = \rho \sum_{j=i}^{N} w_{ij}Y_{jk}^* + S_{ik}\gamma + \varepsilon_{ik} \tag{B.22}$$

with,

$$Y_i = k \quad \text{if} \quad Y_{ik}^* > Y_{il}^*, \forall k \neq l, l = \{1, \dots, L\}$$

where Y_k^* is a latent dependent variable with observable counterpart Y; S_k denotes covariates that are alternative-varying; w_{ij} denotes the spatial weights for observations i and j; ε_k is a vector of IID disturbances for alternative k; and ρ and γ are the model parameters. In this setting, the probability that individual i makes choice k is given by:

$$P_{ik} = \frac{e^{S_{ik}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}} \tag{B.23}$$

These choice probabilities result in the following gradients for the model parameters (γ, ρ).

- Gradient of γ:

$$\frac{\partial P_{ik}}{\partial \gamma} = \frac{\left[(e^{S_{ik}^{**}\gamma})S_{ik}^{**}\right]\left(\sum_l e^{S_{il}^{**}\gamma}\right)}{\left(\sum_l e^{S_{il}^{**}\gamma}\right)^2} - \frac{\left(e^{S_{ik}^{**}\gamma}\right)\left[\sum_l (e^{S_{il}^{**}\gamma}S_{il}^{**})\right]}{\left(\sum_l e^{S_{il}^{**}\gamma}\right)^2}$$

$$= \left(\frac{e^{S_{ik}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}}\right)S_{ik}^{**} - \left(\frac{e^{S_{ik}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}}\right)\sum_l \left(\frac{e^{S_{il}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}}\right)S_{il}^{**}$$

$$= P_{ik}S_{ik}^{**} - P_{ik}\sum_l P_{il}S_{il}^{**}$$

Hence,

$$G_{\gamma_i} = P_{ik}\left[S_{ik}^{**} - \sum_l P_{il}S_{il}^{**}\right] \tag{B.24}$$

- Gradient of ρ:

$$\frac{\partial P_{ik}}{\partial \rho} = \frac{\left[(e^{S_{ik}^{**}\gamma})\left(\left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right)\right]\left(\sum_l e^{S_{il}^{**}\gamma}\right)}{\left(\sum_l e^{S_{il}^{**}\gamma}\right)^2} - \frac{\left(e^{S_{ik}^{**}\gamma}\right)\left[\sum_l\left(e^{S_{il}^{**}\gamma}\left(\left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right)\right]}{\left(\sum_l e^{S_{il}^{**}\gamma}\right)^2}$$

$$= \left(\frac{e^{S_{ik}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}}\right)\left(\left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right) - \left(\frac{e^{S_{ik}^{**}\gamma}}{\sum_l e^{S_{il}^{**}\gamma}}\right)\sum_l \left(\frac{e^{S_{il}^{**}\gamma}}{\sum_l e^{S_{ik}^{**}\gamma}}\right)\left(\left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)$$

$$= P_{ik}\left(\left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right) - P_{ik}\sum_{l} P_{il}\left(\left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)$$

$$= P_{ik}\left[\left(\left(\frac{\partial S_{ik}^{**}}{\partial \rho}\right)\gamma\right) - \sum_{l} P_{il}\left(\left(\frac{\partial S_{il}^{**}}{\partial \rho}\right)\gamma\right)\right]$$

From (B.13): $\frac{\partial S_{ik}^{**}}{\partial \rho} = K_{ik} - \frac{S_{ik}^{**}}{\sigma_i^2}\Lambda_{ii}$ and $\frac{\partial S_{il}^{**}}{\partial \rho} = K_{il} - \frac{S_{il}^{**}}{\sigma_i^2}\Lambda_{ii}$. Hence, the gradient term for ρ is:

$$G_{\rho_i} = P_{ik}\left[\left(K_{ik}\gamma - \frac{S_{ik}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right) - \sum_{l} P_{il}\left(K_{il}\gamma - \frac{S_{il}^{**}\gamma}{\sigma_i^2}\Lambda_{ii}\right)\right] \tag{B.25}$$

where $K = (I - \rho W)^{-1}W(I - \rho W)^{-1}S^*$ and $\Lambda = (I - \rho W)^{-1}W(I - \rho W)^{-1}(I - \rho W)^{-1}$.

At the linearization point ($\rho = 0$), $S_{ik}^{**} = S_{ik}^{*} = S_{ik}$, $S_{il}^{**} = S_{il}^{*} = S_{il}$, and $\Lambda_{ii} = w_{ii} = 0$, thus the terms containing Λ_{ii} in (B.25) vanish. However, the gradient of ρ does not vanish because of the K terms, which reduce to WS. Thus, this model can also be linearized around a point of parameter values which sets $\rho = 0$ because the parameter ρ remains identified from the corresponding gradient. At the linearization point, the gradients for the SCL model simplify to:

$$\begin{aligned} G_{\gamma_i} &= P_{ik}\left[S_{ik} - \sum_{l} P_{il}S_{il}\right] \\ G_{\rho_i} &= P_{ik}\left[(WS)_{ik}\gamma - \sum_{l} P_{il}\,(WS)_{il}\gamma\right] \end{aligned} \tag{B.26}$$

Spatial Ordered Logit (SOL) Model

Consider a SAL specification of an ordered discrete-choice model:

$$Y_i^* = \rho\sum_{j=1}^{N} w_{ij}\,Y_j^* + X_i\beta + \varepsilon_i \tag{B.27}$$

with,

$$Y_i = l \; if \; \alpha_{l-1} < Y_i^* \leq \alpha_l \,, \forall l = \{1, \dots, L\}$$

where Y^* denotes a latent dependent variable with observable counterpart Y; X denotes a matrix of explanatory variables; w_{ij} denotes the spatial weights relating observations i and j; ρ is the spatial autoregressive parameter; β is a vector of model parameters; and ε is a vector of disturbances. The natural ordering of the alternatives is a result of the latent variable falling into various mutually exclusive and exhaustive ranges given by the threshold parameters α_{l-1} and α_l, with $\alpha_0 = -\infty$ and $\alpha_L = \infty$. The dependent variable Y_i can be defined in a binary form as: $d_{il} = I(Y_i = l)$.

The reduced form of the model is given by:

$$Y_i^* = \sum_{j=1}^{N} \psi_{ij} X_j \beta + \sum_{j=1}^{N} \psi_{ij}\, \varepsilon_j \tag{B.28}$$

with, $$Y_i = l \; if \; \alpha_{l-1} < Y_i^* \leq \alpha_l \,, \forall l = \{1, \dots, L\}$$

where ψ_{ij} are the (i,j) elements of the spatial matrix $(I - \rho W)^{-1}$. The reduced model can be written in matrix notation as follows:

$$Y^* = (I - \rho W)^{-1} X\beta + e, \quad e = (I - \rho W)^{-1}\varepsilon \tag{B.29}$$

with resulting covariance matrix: $V(e) \propto [(I - \rho W)'(I - \rho W)]^{-1}$. This covariance structure also implies autocorrelated and heteroskedastic disturbances. Denote by σ_i^2 the variances of the errors given by the diagonal elements of the covariance matrix $V(e)$. The model can be normalized for heteroskedastic variances as follows:

$$\frac{Y_i^*}{\sigma_i} = \frac{\sum_j \psi_{ij}\, X_j \beta}{\sigma_i} + \frac{\sum_j \psi_{ij}\, \varepsilon_j}{\sigma_i} \tag{B.30}$$

with, $$Y_i = l \;\; \text{if} \;\; \frac{\alpha_{l-1}}{\sigma_i} < \frac{Y_i^*}{\sigma_i} \leq \frac{\alpha_l}{\sigma_i}$$

The choice probabilities for this model are given by:

$$P_{il} = P(Y_i = l) = P[I_i^* \alpha_{l-1} < X_i^{**}\beta + \varepsilon_i^{**} \leq I_i^* \alpha_l] = P[\, I_i^* \alpha_{l-1} - X_i^{**}\beta < \varepsilon_i^{**} \leq I_i^* \alpha_l - X_i^{**}\beta]$$

$$P_{il} = F(I_i^* \alpha_l - X_i^{**}\beta) - F(I_i^* \alpha_{l-1} - X_i^{**}\beta) \tag{B.31}$$

where F is the CDF of $\varepsilon^{**}, X^{**} = (I - \rho W)^{-1} X^*$, $X_i^* = X_i/\sigma_i$; $\varepsilon^{**} = (I - \rho W)^{-1}\varepsilon^*, \varepsilon_i^* = \varepsilon_i/\sigma_i$; and $I_i^* = 1/\sigma_i$ to make the scaling of the threshold parameters more explicit. A logistic distribution for the errors ε^{**} gives rise to the spatial ordered logit (SOL) model with the following choice probabilities:

$$P_{il} = \frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}}{1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}} - \frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}}{1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}} \tag{B.32}$$

The gradients for this model can be obtained by differentiating these choice probabilities with respect to the model parameters $(\alpha_{l-1}, \alpha_l, \beta, \rho)$.

- Gradient of α_{l-1}:

$$\frac{\partial P_{il}}{\partial \alpha_{l-1}} = -\left[\frac{\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)} I_i^*\right]\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)}{\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)} I_i^*\right]}{\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right]$$

$$= -\left[\frac{\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)} + \left(e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2\right] I_i^*}{\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2} - \frac{\left(e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2 I_i^*}{\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right]$$

$$= -\left[\frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}}{\left(1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right] I_i^*$$

Hence,

$$G_{\alpha_{il-1}} = -\left[\frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}}{\left(\left[1 + e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right]^2\right)}\right] I_i^* \tag{B.33}$$

- Gradient of α_l:

$$\frac{\partial P_{il}}{\partial \alpha_l} = \frac{\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)} I_i^*\right]\left(1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)}{\left(1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)} I_i^*\right]}{\left(1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}$$

$$= \frac{\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)} + \left(e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2\right] I_i^*}{\left(1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2} - \frac{\left(e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2 I_i^*}{\left(1 + e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}$$

$$= \left[\frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right] I_i^*$$

Thus,

$$G_{\alpha_{il}} = \left[\frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right] I_i^* \tag{B.34}$$

- Gradient of β:

$$\frac{\partial P_{il}}{\partial \beta} = \frac{\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)}(-X_i^{**})\right](1+e^{(I_i^*\alpha_l - X_i^{**}\beta)})}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)}(-X_i^{**})\right]}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}$$

$$- \frac{\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}(-X_i^{**})\right](1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)})}{\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}(-X_i^{**})\right]}{\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}$$

$$= -\left[\frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right] X_i^{**} + \left[\frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right] X_i^{**}$$

Hence,

$$G_{\beta_i} = \left[\frac{e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2} - \frac{e^{(I_i^*\alpha_l - X_i^{**}\beta)}}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right] X_i^{**} \tag{B.35}$$

- Gradient of ρ:

$$\frac{\partial P_{il}}{\partial \rho} = \left[\frac{\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)}\left(\left(\frac{\partial I_i^*}{\partial \rho}\right)\alpha_l - \left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta\right)\right]\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right.$$

$$\left. - \frac{\left(e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)\left[e^{(I_i^*\alpha_l - X_i^{**}\beta)}\left(\left(\frac{\partial I_i^*}{\partial \rho}\right)\alpha_l - \left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta\right)\right]}{\left(1+e^{(I_i^*\alpha_l - X_i^{**}\beta)}\right)^2}\right]$$

$$- \left[\frac{\left[e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\left(\left(\frac{\partial I_i^*}{\partial \rho}\right)\alpha_{l-1} - \left(\frac{\partial X_i^{**}}{\partial \rho}\right)\beta\right)\right]\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)}{\left(1+e^{(I_i^*\alpha_{l-1} - X_i^{**}\beta)}\right)^2}\right.$$

$$-\frac{\left(e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}\right)\left[e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}\left(\left(\frac{\partial I_i^*}{\partial\rho}\right)\alpha_{l-1}-\left(\frac{\partial X_i^{**}}{\partial\rho}\right)\beta\right)\right]}{\left(1+e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}\right)^2}\Bigg]$$

$$=\left[\frac{e^{\left(I_i^*\alpha_l-X_i^{**}\beta\right)}}{\left(1+e^{\left(I_i^*\alpha_l-X_i^{**}\beta\right)}\right)^2}\right]\left[\left(\frac{\partial I_i^*}{\partial\rho}\right)\alpha_l-\left(\frac{\partial X_i^{**}}{\partial\rho}\right)\beta\right]$$

$$-\left[\frac{e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}}{\left(1+e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}\right)^2}\right]\left[\left(\frac{\partial I_i^*}{\partial\rho}\right)\alpha_{l-1}-\left(\frac{\partial X_i^{**}}{\partial\rho}\right)\beta\right]$$

From (B.13): $\frac{\partial X_i^{**}}{\partial\rho}=H_i-\frac{X_i^{**}}{\sigma_i^2}\Lambda_{ii}$. In addition, note that $I_i^*=1/\sigma_i$ which can be written in terms of the error variances σ_i^2 as follows: $I_i^*=1/\sqrt{\sigma_i^2}$. Hence,

$$\frac{\partial I_i^*}{\partial\rho}=\frac{\partial I_i^*}{\partial\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\rho}=-\frac{1}{2(\sigma_i^2)^{3/2}}\frac{\partial\sigma_i^2}{\partial\rho}=-\frac{I_i^*}{2\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\rho}$$

From (B.11): $\frac{\partial\sigma_i^2}{\partial\rho}=2\Lambda_{ii}$, thus, $\frac{\partial I_i^*}{\partial\rho}=-\frac{I_i^*}{\sigma_i^2}\Lambda_{ii}$. The gradient of ρ is given by:

$$G_{\rho i}=\left[\frac{e^{\left(I_i^*\alpha_l-X_i^{**}\beta\right)}}{\left(1+e^{\left(I_i^*\alpha_l-X_i^{**}\beta\right)}\right)^2}\right]\left[\left(-\frac{I_i^*\alpha_l}{\sigma_i^2}\Lambda_{ii}\right)-\left(H_i\beta-\frac{X_i^{**}\beta}{\sigma_i^2}\Lambda_{ii}\right)\right]$$
$$-\left[\frac{e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}}{\left(1+e^{\left(I_i^*\alpha_{l-1}-X_i^{**}\beta\right)}\right)^2}\right]\left[\left(-\frac{I_i^*\alpha_{l-1}}{\sigma_i^2}\Lambda_{ii}\right)-\left(H_i\beta-\frac{X_i^{**}\beta}{\sigma_i^2}\Lambda_{ii}\right)\right] \tag{B.36}$$

where $H=(I-\rho W)^{-1}W(I-\rho W)^{-1}X^*$ and $\Lambda=(I-\rho W)^{-1}W(I-\rho W)^{-1}(I-\rho W)^{-1}$. At the linearization point ($\rho=0$), $X_i^{**}=X_i^*=X_i$, $I_i^*=1$, and $\Lambda_{ii}=w_{ii}=0$, thus the terms containing Λ_{ii} in (B.36) vanish. The gradient of ρ in (B.36) is nonzero when $\rho=0$ because of the term H which becomes WX, thus this model can be linearized as well around $\rho=0$.

The gradient terms for the linearized SOL model are:

$$G_{\alpha_{il-1}} = -\left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2}\right]$$

$$G_{\alpha_{li}} = \left[\frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i\beta)})^2}\right]$$

$$G_{\beta_i} = \left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2} - \frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i\beta)})^2}\right] X_i \tag{B.37}$$

$$G_{\rho i} = \left[\frac{e^{(\alpha_{l-1}-X_i\beta)}}{(1+e^{(\alpha_{l-1}-X_i\beta)})^2} - \frac{e^{(\alpha_{l}-X_i\beta)}}{(1+e^{(\alpha_{l}-X_i\beta)})^2}\right] (WX)_i\beta$$

APPENDIX C
DESCRIPTION OF MONTE CARLO EXPERIMENTS

Scenario 1: SMNL Model with Equal Choice Probabilities

In this scenario, we consider a model with four choice alternatives, each with the same choice probability ($P_0 = P_1 = P_2 = P_3 = 0.25$). To generate the data for this model we start with the reduced form of the model written in matrix notation as:

$$Y^* = (I - \rho W)^{-1} X\beta + e, \qquad e = (I - \rho W)^{-1}\varepsilon \tag{C.1}$$

with covariance matrix $V(e) = [(I - \rho W)'(I - \rho W)]^{-1}$. We generate a single explanatory variable (X) uniformly distributed in the interval $(-1, 1)$. Next, we specify the spatial weights matrix W as a row-standardized first-order contiguity matrix, which sets the non-zero elements to: $w_{ij} = 0.5 \;\; if \; |i - j| = 1$ with endpoints $w_{1,2} = w_{N,N-1} = 1$, and set a value of ρ between 0 and 0.9, varying it in increments of 0.1. Having specified W and a value for ρ, the explanatory variable is first transformed to obtain X^* as $X_i^* = X_i/\sigma_i$ where σ_i is the square root of the error variances given by the diagonal of $V(e)$. A second transformation follows to obtain X^{**} as $X^{**} = (I - \rho W)^{-1} X^*$. For simplicity, each of the parameters β_k is set equal to 1. Equal parameter values for the different choice alternatives and a fixed X (thus, a fixed X^{**}) ensure equal choice probabilities. Subsequently, the simulated probabilities are obtained as:

$$P_{ik} = \frac{\mathrm{e}^{X_i^{**}\beta_k}}{\sum_l \mathrm{e}^{X_i^{**}\beta_l}}. \tag{C.2}$$

To generate the observed individual choices based on the simulated probabilities, we generate a uniform $(0, 1)$ random variable (u) and set $d_{ik} = 1$ for $k = l$ if:

$$\sum_{k=0}^{l-1} P_{ik} < u < \sum_{k=0}^{l} P_{ik}, \text{ with } P_{i0} = 0 \tag{C.3}$$

Scenario 2: SMNL Model with Different Choice Probabilities

In this scenario, we still consider four choice alternatives, but now each alternative has a different choice probability. We are interested in alternatives with the following choice probabilities: $P_0 = 0.10,\ P_1 = 0.50,\ P_2 = 0.25,\ P_3 = 0.15$. To generate this data, we start with the reduced form of the model in (C.1) and generate a single explanatory variable (X) uniformly distributed in the interval $(0.5, 1.5)$. Next, we set the desired choice probabilities and solve for the parameters β_k that correspond to those probabilities. Let $l = 1,2,3,4$ be the choice alternatives and i be implicit. Then, the SMNL model can be written as:

$$Y_l^* = X\beta_l + \varepsilon_l\ ,\ \varepsilon_l \sim IID\ logistic \tag{C.4}$$

Given the model in (C.4), the choice probabilities are given by:

$$P_k = \frac{\mathrm{e}^{X^{**}\beta_k}}{\sum_l \mathrm{e}^{X^{**}\beta_l}}. \quad k, l \in \{1,2,3,4\} \tag{C.5}$$

With a fixed X, we can fix P_k by solving for β_k. Starting out with P_1:

$$P_1 = \frac{\mathrm{e}^{X^{**}\beta_1}}{\sum_{l=1}^{4} \mathrm{e}^{X^{**}\beta_l}}$$

$$\mathrm{e}^{X^{**}\beta_1} - P_1\left(\mathrm{e}^{X^{**}\beta_1}\right) = P_1 \sum_{l\neq 1} \mathrm{e}^{X^{**}\beta_l}$$

$$\mathrm{e}^{X^{**}\beta_1} = \left(\frac{P_1}{1-P_1}\right)\sum_{l\neq 1} \mathrm{e}^{X^{**}\beta_l}$$

By symmetry,

$$\mathrm{e}^{X^{**}\beta_2} = \left(\frac{P_2}{1-P_2}\right)\sum_{l\neq 2} \mathrm{e}^{X^{**}\beta_l}$$

$$\mathrm{e}^{X^{**}\beta_3} = \left(\frac{P_3}{1-P_3}\right)\sum_{l\neq 3} \mathrm{e}^{X^{**}\beta_l}$$

$$\mathrm{e}^{X^{**}\beta_4} = \left(\frac{P_4}{1-P_4}\right)\sum_{l\neq 4} \mathrm{e}^{X^{**}\beta_l}$$

Normalizing $\beta_1 = 0$:

$$e^{X^{**}\beta_2} = \left(\frac{P_2}{1-P_2}\right)\left(1 + e^{X^{**}\beta_3} + e^{X^{**}\beta_4}\right) \tag{C.6}$$

$$e^{X^{**}\beta_3} = \left(\frac{P_3}{1-P_3}\right)\left(1 + e^{X^{**}\beta_2} + e^{X^{**}\beta_4}\right) \tag{C.7}$$

$$e^{X^{**}\beta_4} = \left(\frac{P_4}{1-P_4}\right)\left(1 + e^{X^{**}\beta_2} + e^{X^{**}\beta_3}\right) \tag{C.8}$$

The solution to β_2 , β_3 , β_4 should also satisfy:

$$1 = \left(\frac{P_1}{1-P_1}\right)\left(e^{X^{**}\beta_2} + e^{X^{**}\beta_3} + e^{X^{**}\beta_4}\right) \tag{C.9}$$

Substituting (C.6) into (C.7):

$$e^{X^{**}\beta_3} = \left(\frac{P_3}{1-P_3}\right)\left[1 + \left(\left(\frac{P_2}{1-P_2}\right)\left(1 + e^{X^{**}\beta_3} + e^{X^{**}\beta_4}\right)\right) + e^{X^{**}\beta_4}\right]$$

$$= \left(\frac{P_3}{1-P_3}\right) + \left(\frac{P_3}{1-P_3}\right)\left(\left(\frac{P_2}{1-P_2}\right)\left(1 + e^{X^{**}\beta_3} + e^{X^{**}\beta_4}\right)\right)$$

$$+ \left(\frac{P_3}{1-P_3}\right) e^{X^{**}\beta_4}$$

$$= \left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2} + 1\right) + \left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2}\right) e^{X^{**}\beta_3}$$

$$+ \left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2} + 1\right) e^{X^{**}\beta_4}$$

$$e^{X^{**}\beta_3} = \frac{\left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2} + 1\right)}{\left[1 - \left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2}\right)\right]} e^{X^{**}\beta_4} + \frac{\left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2} + 1\right)}{\left[1 - \left(\frac{P_3}{1-P_3}\right)\left(\frac{P_2}{1-P_2}\right)\right]} \tag{C.10}$$

Similarly, substituting (C.6) into (C.8):

$$e^{X^{**}\beta_4} = \left(\frac{P_4}{1-P_4}\right)\left[1 + \left(\left(\frac{P_2}{1-P_2}\right)\left(1 + e^{X^{**}\beta_3} + e^{X^{**}\beta_4}\right)\right) + e^{X^{**}\beta_3}\right]$$

$$= \left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}+1\right)+\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}+1\right)\mathrm{e}^{X^{**}\beta_3}$$

$$+\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}\right)\mathrm{e}^{X^{**}\beta_4}$$

$$\mathrm{e}^{X^{**}\beta_4} = \frac{\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}+1\right)}{\left[1-\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}\right)\right]}\mathrm{e}^{X^{**}\beta_3}+\frac{\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}+1\right)}{\left[1-\left(\frac{P_4}{1-P_4}\right)\left(\frac{P_2}{1-P_2}\right)\right]} \tag{C.11}$$

To simplify the notation, let:

$$a=\left(\frac{P_1}{1-P_1}\right), b=\left(\frac{P_2}{1-P_2}\right), c=\left(\frac{P_3}{1-P_3}\right), d=\left(\frac{P_4}{1-P_4}\right)$$

Then,

$$\mathrm{e}^{X^{**}\beta_3} = \frac{c(b+1)}{[1-(c)(b)]}\mathrm{e}^{X^{**}\beta_4}+\frac{c(b+1)}{[1-(c)(b)]} \tag{C.12}$$

$$\mathrm{e}^{X^{**}\beta_4} = \frac{d(b+1)}{[1-(d)(b)]}\mathrm{e}^{X^{**}\beta_3}+\frac{d(b+1)}{[1-(d)(b)]} \tag{C.13}$$

Substituting (C.12) into (C.13) and solving for β_4:

$$\mathrm{e}^{X^{**}\beta_4} = \frac{d(b+1)}{[1-(d)(b)]}\left[\frac{c(b+1)}{[1-(c)(b)]}\mathrm{e}^{X^{**}\beta_4}+\frac{c(b+1)}{[1-(c)(b)]}\right]+\frac{d(b+1)}{[1-(d)(b)]}$$

$$= \left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}\right)\mathrm{e}^{X^{**}\beta_4}+\left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}\right)$$

$$+\left(\frac{d(b+1)}{[1-(d)(b)]}\right)$$

$$\left[1-\left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}\right)\right]\mathrm{e}^{X^{**}\beta_4} = \left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}+1\right)$$

$$\mathrm{e}^{X^{**}\beta_4} = \frac{\left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}+1\right)}{\left[1-\left(\frac{d(b+1)}{[1-(d)(b)]}\right)\left(\frac{c(b+1)}{[1-(c)(b)]}\right)\right]} = A \tag{C.14}$$

Thus,

$$\beta_4 = \frac{1}{X^{**}} \ln(A) \tag{C.15}$$

Plugging (C.14) into (C.12) and solving for β_3:

$$\mathrm{e}^{X^{**}\beta_3} = \frac{c(b+1)}{[1-(c)(b)]} A + \frac{c(b+1)}{[1-(c)(b)]} = \frac{c(b+1)}{[1-(c)(b)]}(A+1)$$

$$\mathrm{e}^{X^{**}\beta_3} = \frac{c(b+1)}{[1-(c)(b)]}(A+1) = B \tag{C.16}$$

Hence,

$$\beta_3 = \frac{1}{X^{**}} \ln(B) \tag{C.17}$$

Finally, plugging (C.14) and (C.16) into (C.6) and solving for β_2:

$$\mathrm{e}^{X^{**}\beta_2} = b(1 + \mathrm{B} + \mathrm{A}) = C \tag{C.18}$$

Thus,

$$\beta_2 = \frac{1}{X^{**}} \ln(C) \tag{C.19}$$

It becomes evident from (C.15), (C.17), and (C.19) that the parameters β_k are now a function of X^{**} – this is the reason for changing the interval of (X) to ensure a mean of 1 rather than 0 – and as a result a function of ρ, hence denoted by $\beta_k(\rho)$. As in the first scenario, we set the value of ρ between 0 and 0.9, varying it in increments of 0.1 and specify W as a first-order contiguity matrix. The explanatory variable is transformed to obtain X^* and X^{**} as described above and the observed individual choices are assigned based on the simulated probabilities given by (C.2) using the assignment rule in (C.3).

LIST OF REFERENCES

Ahearn, M.C., Alig,R.J, 2006. A discussion of recent land-use trends. In: Bell, K.P., Boyle, K.J., Rubin, J. (Eds.), Economics of Rural Land-Use Change. Ashgate Studies in Environmental and Natural Resource Economics, pp. 11-25.

Ahearn, M., Perry, J., El-Osta, H., 1993. The economic well-being of farm operator households. U.S. Department of Agriculture, Economic Research Service Agricultural Economic Report No. 666, Washington, D.C.

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669-679.

Alberti, M., 2005. The effects of urban patterns on ecosystem function. International Regional Science Review 28(2), 168-192.

Alig, R.J., Ahearn, M.C., 2006. Effects of policy and technological change on land use. In: Bell, K.P., Boyle, K.J., Rubin, J. (Eds.), Economics of Rural Land-Use Change. Ashgate Studies in Environmental and Natural Resource Economics, pp. 27-40.

Amemiya, T., 1985. Advanced Econometrics. Cambridge, MA: Harvard University Press.

Amrhein, C., Guevara, J., Griffith, D., 1983. The effects of random Thiessen structure and random processes of the measurement of spatial autocorrelation. Modeling and Simulation 14, 585-589.

Anas, A., Arnott, R., Small, K., 1998. Urban spatial structure. Journal of Economic Literature 36, 1426-1464.

Anas, A., 2001. The costs and benefits of fragmented metropolitan governance and the new regional policies. Planning and Markets 2 (1). Retrieved from http://www-pam.usc.edu.

Andrews, D., 1986. A note on the unbiasedness of feasible GLS, quasi-maximum likelihood, robust, adaptive, and spectral estimators of the linear model. Econometrica 54, 687-698.

Angresti, A., 1996. An Introduction to Categorical Data Analysis. New York: John Wiley.

Anselin, L., 1980. Estimation Methods for Spatial Autoregressive Structures. Regional Science Dissertation and Monograph Series 8, Cornell University, Ithaca, NY.

Anselin, L., 1981. Small sample properties of estimators for the linear model with a spatial autoregressive structure in disturbance. Modeling and Simulation 12, 899-904.

Anselin, L., 1984. Specification tests on the structure of interaction in spatial econometric models. Papers in Regional Science 54 (1), 165-182.

Anselin, L., 1988a. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin, L., 1988b. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis 20, 1-17.

Anselin, L., 1993. Discrete space autoregressive models. In: Goodchild, M.F., Parks, B.O., Steyaert, T. (Eds.), Environmental Modelling with GIS. Oxford: Oxford University Press, pp. 454-469.

Anselin, L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models. Agricultural Economics 17, 247-267.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review 26, 153-66.

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D. (Eds.), Handbook of Applied Economic Statistics. New York: Marcel Dekker, pp. 237-289.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26, 77-104.

Anselin,L., Can, A., 1996. Spatial effects in models of mortgage origination. Paper presented at the Mid Year AREUEA Conference, Washington D.C., May 28-29.

Arnott, R.J., Lewis, F.D., 1979. The transition of land to urban use. The Journal of Political Economy 87(1), 161-169.

Armsworth, P.R., Kendall, B.E., Davis, F.W., 2004. An introduction to biodiversity concepts for environmental economists. Resource and Energy Economics 26, 115-136.

Attkisson, L.R., 2009. Putting a stop to sprawl: state intervention as a tool for growth management. Vanderbilt Law Review 62(3), 979-1015.

Avery, R.B., Hansen, L.P., Holtz, V.J., 1983. Multiperiod probit models and orthogonality condition estimation. International Economic Review 24, 21-35.

Batchis, W., 2010. Enabling urban sprawl: revisiting the Supreme Court's seminal zoning decision Euclid vs. Amber in the 21st century. Virginia Journal of Social Policy and the Law 17, 373-403.

Bates, C., White, H., 1985. A unified theory of consistent estimation for parametric models. Econometric Theory 1, 151-178.

Belfield, C.R., Levin, H.M., 2002. The effects of competition on educational outcomes: a review of U.S. evidence. National Center for the Study of Privatization in Education, Columbia University. Retrieved from http://www.ncspe.org/publications_files/688_OP35V2.pdf

Bell, K., Bockstael, N., 2000. Applying the generalized methods of moments approach to spatial problems involving micro-level data. Review of Economics and Statistics 82, 72-82.

Bell, K. P., Irwin, E.G., 2002. Spatially explicit micro-level modeling of land use change at the urban rural interface. Agricultural Economics 27, 217-132.

Ben-Akiva, M., McFadden, D., Abe, M., Bockenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramawamy, V., Rao, V., Revelt, D., Steinberg, D., 1997. Modeling methods for discrete choice analysis. Marketing Letters 8(3), 273-286.

Benirschka, M., Binkley, J.K., 1994. Land price volatility in a geographically dispersed market. American Journal of Agricultural Economics 76, 185-195.

Bera, A. K., McKenzie, C.R., 1986. Alternative forms and properties of the score test. Journal of Applied Statistics 13,13-25.

Bera, A.K., Yoon, M.J., 1993. Specification testing with misspecified alternatives. Econometric Theory 9, 649-658.

Beron, K.J., Vijverberg, W.P., 2000. Probit in a spatial context: a Monte Carlo approach. In: Anselin, L., Florax, R. (Eds.), Advances in Spatial Econometrics. Heidelberg: Springer-Verlag.

Beron, K.J., Vijverberg, W.P., 2003. Probit in a spatial context: a Monte Carlo approach. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics. Heidelberg: Springer-Verlag.

Beron, K.J., Vijverberg, W.P., 2004. Probit in a spatial context: a Monte Carlo analysis. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer-Verlag, pp. 169-195.

Besley, T., Case, A., 1995. Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. American Economic Review 85, 25-45.

Bier, T., 2001. Urban sprawl and decline. Prospects for change. Public Works Management and Policy 6(2), 83-87.

Blommestein, H., 1983. Specification and estimation of spatial econometric models: a discussion of alternative strategies for spatial economic modeling. Regional Science and Urban Economics 13, 250-271.

Blommestein, H., 1985. Elimination of circular routes in spatial dynamic regression equations. Regional Science and Urban Economics 15, 121-130.

Bodson, P., Peeters, D., 1975. Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation: an application to a Belgian labor-demand function. Environment and Planning A 7, 455-472.

Bockstael, N.E., 1996. Modeling economics and ecology: the importance of a spatial perspective. American Journal of Agricultural Economics 78(5), 1168-1180.

Bockstael, N.E., Bell, K., 1997. Land use patterns with water quality: the effect of differential land management controls. In: Just, R., Netanyahu, S. (Eds.), International Water and Resource Economics Consortium, Conflict and Cooperation on Transboundary Water Resources. Kluwer Academic Publishers, Norwell, MA, pp. 169-191.

Borooah, V.K., 2002. Logit and Probit: Ordered and Multinomial Models. Sage University Paper Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

Bowden, R., Turkington, D., 1984. Instrumental Variables. London: Cambridge University Press.

Brandsma, A.S., Ketellapper, R.H., 1979. A biparametric approach to spatial autocorrelation. Environment and Planning A 11, 51-58.

Brett, C., Pinkse, J., 1997. Those taxes are all over the map: a test for spatial dependence of municipal tax rates in British Columbia. International Regional Science Review 20, 131-151.

Brett, C., Pinkse, J., 2000. The determinants of municipal tax rates in British Columbia. Canadian Journal of Economics 33, 695-714.

Brown, D.G., Johnson, K.M., Loveland, T.R., Theobald, D.M., 2005. Rural land-use trend in the conterminous United States, 1950-2000. Ecological Applications 15(6), 1851-1963.

Brueckner, J.K., 1998. Testing for strategic interaction among local governments: the case of growth controls. Journal of Urban Economics 44, 438-467.

Brueckner, J.K., 2000. Urban sprawl: diagnosis and remedies. International Regional Science Review 23(2), 160-171.

Brueckner, J.K., Saavedra, L.A., 2001. Do local governments engage in strategic property tax competition? National Tax Journal 54(2), 203-229.

Buddin, R.J., Cordes, J.J., Kirby, S.N., 1998. School choice in California: who chooses private schools? Journal of Urban Economics 44, 110-134.

Burchell, R.W., Shad, N.A., Listokin, D., Phillips, H., Downs, A., Seksin, S., Davis, J., Moore, T., Helton, D., Gall, M., 1998. The costs of sprawl – revisited. Transportation Research Board, National Research Council. National Academy Press, Washington, DC.

Burridge, P., 1980. On the Cliff-Ord test for spatial autocorrelation. Journal of the Royal Statistical Society B 42, 107-108.

Cameron, A., Trivedi, P., 2005. Microeconometrics: Methods and Applications. Cambridge: Cambridge University Press.

Can, A., 1992. Specification and estimation of hedonic housing price models. Regional Science and Urban Economics 22, 453-474.

Can, A., Megbolugbe, I.F., 1997. Spatial dependence and house price index construction. Journal of Real Estate Finance and Economics 14, 203-222.

Cappoza, D.R., Helsley, R.W., 1989. The fundamentals of land prices and urban growth. Journal of Urban Economics 26, 295-306.

Cappozza, D.R., Helsley, R.W., Mills, E.S., 1986. Urban growth and the price of land. Proceedings, Administrative Science Association of Canada, 1986.

Carlson, D., Lavery, L., Witte, J.F., 2011. The determinants of interdistrict open enrollment flows. Evidence from two states. Educational Evaluation and Policy Analysis 33(1), 76-94.

Carrión-Flores, C.E., Irwin, E.G., 2004. Determinants of residential land use conversion and sprawl at the rural-urban fringe. American Journal of Agricultural Economics 86(4), 889-904.

Case, A.C, 1987. On the use of spatial autoregressive models in demand analysis. Discussion Paper 135, Research Program in Development Studies, Woodrow Wilson School, Princeton University.

Case, A.C, 1991. Spatial patterns in household demand. Econometrica 59, 953-965.

Case, A.C, 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22, 491-508.

Case, A.C., Rosen, H.S., Hines, J.R., 1993. Budget spillovers and fiscal policy independence: evidence from the states. Journal of Public Economics 52, 285-307.

Chakir, R., Parent, O., 2009. Determinants of land use changes: a spatial multinomial probit approach. Papers in Regional Science 88(2), 327-344.

Chubb, J., Moe, T., 1988. Politics, markets, and the organization of schools. American Political Science Review 82(4), 1065-1087.

Chubb, J., Moe, T., 1990. Politics, Markets, and America's Schools. Washington D.C: The Brookings Institution.

Cliff, A., Ord, J.K., 1972. Testing for spatial autocorrelation among regression residuals. Geographic Analysis 4, 267-284.

Cliff, A., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Cliff, A., Ord, J.K., 1981. Spatial Processes, Models and Applications. Pion, London.

Coleman, J.S., Schiller, K., Schneider, B.,1993. Parent choice and inequality. In: Schneider, B., Coleman, J. (Eds.), Parents, Their Children, and Schools. Westview Press, Boulder, CO, pp. 147-182.

Collins, J.P., Kinzig, A., Grimm, N.B., Fagan, W.F., Hope, D., Wu, J.G., Borer, E.T., 2000. A new urban ecology. American Scientist 88(5), 416-425.

Cook, D., Pocock, S., 1983. Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. Biometrics 39, 361-371.

Cookson, P.W., Jr., 1994. School Choice: The Struggle for the Soul of American Education. New Haven, NJ: Yale University Press.

Crain, R.L, Heebner, A.L., Yiu-Pong S., 1992. The effectiveness of New York City's career magnet schools: an evaluation of ninth-grade performance using an experimental design. National Center for Research in Vocational Education, Berkley, CA.

Crepage, R. A., 1999. School Choice and Ohio's Interdistrict Open Enrollment Policy. Doctoral Dissertation. Youngstown State University.

Cullen, J.B., Jacob, B.A., Levitt, S.D., 2005. The impact of school choice on student outcomes: an analysis of the Chicago public schools. Journal of Public Economics 89, 729-760.

Dacey, M., 1968. A review of measures of contiguity for two and K-color maps. In: Berry, B., Marble, D. (Eds.), Spatial Analysis: A Reader in Statistical Geography. Englewood Cliffs, N.J.: Prentice Hall, pp. 479-495.

Daly, A., Zachary, S., 1979. Improved multiple choice models. In: Hensher, D., Dalvi, Q. (Eds.), Measuring the Determinants of Mode Choice. London, Teakfield.

Dempster, A.P., Laird, N.M., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1-38.

Domencich, T.A., McFadden, D., 1996. Urban Travel Demand: A Behavioral Analysis. Amsterdam: North Holland.

Doreian, P., 1980. Linear models with spatially distributed data, spatial disturbances, or spatial effects. Sociological Methods and Research 9, 29-60.

Doreian, P., 1982. Maximum likelihood methods for linear models: spatial effect and spatial disturbance terms. Sociological Methods and Research 13, 243-269.

Duax, T., 1988. The Impact of Magnet Schools on a Predominantly Black Community. University of Wisconsin, Madison, WI.

Elam, S.M., Rose, L.C., Gallup, A.M., 1997. The 29th annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan 79(1), 41-58.

Ewing, R.,1997. Is Los Angeles-style sprawl desirable? Journal of American Planning Association 63(1), 49-80.

Farrell, R.B., 1994. A Policy Analysis of Interdistrict Open Enrollment in Ohio: Expectations and Outcomes. Doctoral Dissertation, University of Cincinnati.

Feitelson, E., 1993. The spatial effects of land use regulations: a missing link in the growth control evaluations. Journal of American Planning Association 59(4), 461-472.

Fischel, W.A., 1990. Do growth controls matter? A review of the empirical evidence on the effectiveness and efficiency of local government and land use regulation. Lincoln Institute of Land Policy, Cambridge.

Fleming, M.M., 2004. Techniques for estimating spatially dependent discrete choice models. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics: Methodology, Tools, and Applications. New York: Springer-Verlag, pp. 145-168.

Florax, R., Rey, S., 1995. The impacts of misspecified spatial interaction in linear regression models. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 111-135.

Flores-Lagunes, A., Schiner, K.E., 2010. Estimation of sample selection models with spatial dependence. Journal of Applied Econometrics. Published online in Wiley Interscience. DOI: 10.1002/jae.1189.

Flowerdew, R., Green, M., 1989. Statistical methods for inference between incompatible zonal systems. In: Goodchild, M., Gopal, S. (Eds.), Accuracy of Spatial Databases. New York: Taylor and Francis, pp. 239-247.

Fowler, F.C., 1996. Participation in Ohio's interdistrict open enrollment option: exploring the supply-side of choice. Educational Policy 10(4), 518-536.

Frank, J.E., 1989. The costs of alternative development patterns: a review of literature. Urban Land Institute, Washington, DC.

Fredriksson, P.G., Millimet, D.L., 2002. Strategic interaction and the determinants of environmental policy across U.S. states. Journal of Urban Economics 51, 101-122.

Friedman, M., 1955. The role of government in education. In: Solo, R. (Ed.), Economics and Public Interest. New Brunswick: Rutgers University Press, pp. 123-144.

Fujita, M., 1982. Spatial patterns of residential development. Journal of Urban Economics 12, 22-52.

Gardner, B.D., 1977. The economics of agricultural land preservation. American Journal of Agricultural Economics 59(5), 1027-1036.

Gallant, A., Jorgenson, D., 1979. Statistical inference for a system of non-linear implicit equations in the context of instrumental variables estimation. Journal of Econometrics 11, 275-302.

Geary, R., 1954. The contiguity ratio and statistical mapping. The Incorporated Statistician 5, 115-145.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721-741.

Getis, A., 1995. Spatial filtering in a regression framework: examples using data on urban crime, regional inequality, and government expenditures. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 172-185.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317-1340.

Godfrey, L., 1981. On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis. Econometrica 49,1443-1455.

Goldhaber, D.D., Brewer, D.J., Eide, E.R.,Rees, D.I., 1999. Testing for sample selection in the Milwaukee school choice experiment. Economics of Education Review 18(2), 259-267.

Gordon, P., Richardson, H., 1997. Are compact cities a desirable planning goal? Journal of the American Planning Association 63(1), 95-106.

Green, J.P., Peterson, P.E., Du, J., 1999. The effectiveness of school choice: the Milwaukee experiment. Education and Urban Society 31(2), 190-213.

Greene, W.H., 2000. Econometric Analysis. Englewood Cliffs, NJ: Prentice Hall (4th ed.)

Greene, W.H., 1999. Econometric Analysis. New York: Macmillan Publishing Company.

Griffith, D., 1990. A numerical simplification for estimating parameters of spatially autoregressive models. In: Griffith, D. (Ed.), Spatial Statistics: Past, Present, and Future. Ann Arbor, Michigan: Institute of Mathematical Geography, pp 183-1897.

Griffith, D.,1992. Simplifying the normalizing factor in spatial autoregressions for irregular lattices. Papers in Regional Science 71(1), 71-86.

Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.H., Wu, J., Bai, X., Briggs, J.M., 2008. Global change and the ecology of cities. Science 319(5864), 756- 760.

Haining, R.P., 1978. The moving average model for spatial interaction. Transactions, Institute of British Geographers 3, 202-225.

Haining, R.P., 1988. Estimating spatial means with an application to remotely sensed data. Communications in Statistics: Theory and Methods 17, 573-597.

Haining, R.P.,1990. Spatial Data Analysis in the Social and Environmental Science. Cambridge University Press.

Hajivassiliou, V. A., 1990. Smooth estimation of panel data LDV models. Working paper, Department of Economics, Yale University, New Haven, CT.

Hajivassiliou, V. A., Ruud, P.A., 1994. Classical estimation methods for LDV models using simulation. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics. Amsterdam: North Holland, Volume 4, pp. 2383-2441.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

Hansen, A.J, Knight, R.L., Marzluff, J.M., Powell, S., Brown, K., Gude, P.H., Jones, K., 2005. Effects of exurban development on biodiversity: patterns, mechanisms, and research needs. Ecological Applications 15(6), 1893-1905.

Hausman, J., McFadden, D., 1984. Specification tests for the multinomial logit model. Econometrica 52(5), 1219-1240.

Heavner, B., 2000. Paving the way. How highway construction has contributed to sprawl in Maryland. MaryPIRG Foundation, November 2000.

Heijmans, R., Magnus, J., 1986a. On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. Statistica Neerlandica 40, 169-188.

Heijmans, R., Magnus, J., 1986b. Consistent maximum-likelihood estimation with dependent observations: the general (non-normal) case and the normal case. Journal of Econometrics 32, 253-285.

Heijmans, R., Magnus, J., 1986c. Asymptotic normality of maximum likelihood estimators obtained from normally distributed but dependent observations. Econometric Theory 2, 374-412.

Heimlich, R.E., Anderson, W.D., 2001. Development at the urban fringe and beyond: impacts on agriculture and rural land. Agricultural Economic Report No.803, U.S. Department of Agriculture, Economic Research Service: Washington, DC.

Henig, J.R., 1994. Rethinking School Choice: Limits of the Market. Princeton, NJ: Princeton University Press.

Hepple, L., 1976. A maximum likelihood model for econometric estimation with spatial series. In: Masser, I. (Ed.), Theory and Practice in Regional Science, London Papers in Regional Science 6. Pion, London, pp. 90-104.

Hepple, L., 1995a. Bayesian techniques in spatial and network econometrics.1: model comparison and posterior odds. Environment and Planning A 27, 447-469.

Hepple, L., 1995b. Bayesian techniques in spatial and network econometrics. 2: computational methods and algorithms. Environment and Planning A 27, 615-644.

Hepple, L., 2000. Spatial autocorrelation. In: Johnson, R., Gregory, D., Pratt, G., Watts, M. (Eds.), The Dictionary of Human Geography. Oxford: Blackwell Publishers.

Hordijk, L., Paelinck, J., 1976. Some principles and results in spatial econometrics. Recherches Economiques de Louvain 42, 175-197.

Horowitz, J.L., Bolduc, D., Divakar, S., Geweke, J., Gonul, F., Hajivassiliou, V., Koppelman, F.S., Keane, M., Matzkin, R., Rossi, P., Ruud, P., 1994. Advances in random utility models. Marketing Letters 5, 311-322.

Hoxby, C.M., 2000. Does competition among public schools benefit students and taxpayers? American Economic Review 90, 1209-1238.

Hoxby, C.M., 2003. School choice and school productivity (or, could school choice be a rising tide that lifts all boats?). In: Hoxby, C.M. (Ed.), The Economics of School Choice. Chicago: University of Chicago Press, pp. 287-341.

Huang,J.S., 1984. The autoregressive moving average model for spatial analysis. Australian Journal of Statistics 26, 169-178.

Hunt, R.T., 2001. The comprehensive plan: an effective litigation tool in zoning cases. Ohio Municipal Service 13, 65-70.

Irwin, E.G., 2002. The effects of open space on residential property values. Land Economics 78(4), 465-480.

Irwin, E.G., Bockstael, N.E., 2002. Interacting agents, spatial externalities and the endogeneous evolution of land use patterns. Journal of Economic Geography 2(1), 31-54.

Irwin, E.G., Bockstael, N.E., 2004. Analysis of urban land markets and the impact of land market regulation. Regional Science and Urban Economics 34(6), 705-725.

Isard, W.,1969. General Theory: Social, Political, Economic, and Regional, with Particular Reference to Decision-making Analysis. Cambridge: MIT Press.

Johnson, N.L., Kotz, S., 1970. Distributions in Statistics: Continuous Univariate Distributions -1. New York, John Wiley.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lee, T.S., 1985. The Theory and Practice of Econometrics. John Wiley and Sons, New York, NY.

Kakamu, K., Wago, H., 2005. Bayesian spatial panel probit model with an application to business cycle in Japan. Proceedings of International Congress on Modeling and Simulation.

Katz, M.L., Shapiro, C., 1986. Technology adoption in the presence of network externalities. Journal of Political Economy 94, 822-841.

Keane, M.P., 1993. Simulation estimation for panel data models with limited dependent variables. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics. Amsterdam: North Holland, Volume 11, pp. 545-571.

Keane, M.P., 1994. A computationally practical simulator estimator for panel data. Econometrica 62, 95-116.

Kelejian, H., Robinson, D., 1992. Spatial autocorrelation: a new computationally simple test with an application to per capita country police expenditures. Regional Science and Urban Economics 22, 317-331.

Kelejian, H., Robinson, D., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297-312.

Kelejian, H., Robinson, D., 1995. Spatial correlation: a suggested alternative to the autoregressive model. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 75-95.

Kelejian, H., Prucha, I., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99-121.

Kelejian, H., Prucha, I., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509-533.

Kelejian, H., Prucha, I., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219-257.

Klier, T., McMillen, D.P., 2008. Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples. Journal of Business Economics and Statistics 26(4), 460-471.

Kline, J.D., 2000. Comparing states with and without growth management. Analysis based on indicators with policy implications. Comment. Land Use Policy 17, 349-355.

Kline, J.D., Alig., R.J., 1999. Does land use planning slow conversion of forest and farmlands? Growth Change 30, 3-22.

Ladd, H.F., 1992. Population growth, density, and the costs of providing public services. Urban Studies 292, 273-295.

Landis, J., Zhang, M., 1998. The second generation of the California urban futures model. Part 1: model logic and theory. Environment Planning A 25, 657-666.

Lee, L., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252-277.

Lee, L., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22 (4), 307-335.

Lee, L., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6), 1899-1925.

Lee, V.E., Croninger, R.G., Smith, J.B., 1994. Parental choice of schools and social stratification: paradox of Detroit. Educational Evaluation and Policy Analysis 16 (4), 434-457.

LeSage, J.P., 1997a. Regression analysis of spatial data. The Journal of Regional Analysis & Policy 27(2), 83-94.

LeSage, J.P., 1997b. Bayesian estimation of spatial autoregressive models. International Regional Science Review 20, 113-129.

LeSage, J.P., 2000. Bayesian estimation of the limited dependent variable spatial autoregressive models. Geographical Analysis 32, 19-35.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall.

Lewis, D.J., 2010. An economic framework for forecasting land-use and ecosystem change. Resource and Energy Economics 32, 98-116.

Li, M., 2010. Three Essays on Land-use Change, Carbon Sequestration, and Emissions in China. Doctoral Dissertation, Oregon State University.

Lieberman, M.L., 1990. Public School Choice: Current Issues, Future Prospects. Lancaster PA: Technomics.

Lubowski, R.N., Plantiga, A.J., Stavins, R.N., 2008. What drives land-use change in the United States? A national analysis of landowners decision. NBER Working Paper Series, Working Paper 13572.

Luce, R.D., 1959. Individual Choice Behavior: A Theoretical Analysis. NewYork, John Wiley.

Luce, D., Suppes, P., 1965. Preferences, utility, and subjective probability. In: Luce, R., Bush, R., Galanter, E. (Eds.), Handbook of Mathematical Psychology. John Wiley and Sons, New York, pp. 249-410.

Lutz, S., 1996. The impact of school choice in the United States and the Netherlands on ethnic segregation, and equal educational opportunity. Equity and Excellence in Education 29(3), 48-54.

Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Economics. Cambridge, UK, Cambridge University Press.

Maestas, J.D., 2007. Effects of exurban development on wildlife and plant communities. U.S. Department of Agriculture, Natural Resources Conservation Service. Technical Note No. 75, Washington, DC.

Magnus, J., 1978. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. Journal of Econometrics 7, 281-312. (Corrigenda, Journal of Econometrics, 10, 261)

Marschak, J., 1960. Binary choice constraints on random utility indicators. Arrow, K., (Ed.), Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford Univ. Press: Stanford.

McCulloch, R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207-240.

McCulloch, R., Polson, N., Rossi, P., 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. Journal of Econometrics 99, 173-193.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarambka, P. (Ed.), Frontiers in Econometrics. New York, New York: Academic Press, pp. 105-142 .

McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. Cambridge, MA, MIT Press, pp. 198-272.

McFadden, D., 1984. Econometric analysis of qualitative response models. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics. Amsterdam: North Holland, Volume 2, pp. 1395-1457.

McFadden, D., 1987. Regression-based specification tests for the multinomial logit model. Journal of Econometrics 34, 63-82.

McFadden, D., 2001. Economic choices. American Economic Review 91, 351-378.

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

McMillen, D.P., 1989. An empirical model of urban fringe land use. Land Economics 65(2), 138-145.

McMillen, D.P., 1992. Probit with spatial autocorrelation. Journal of Regional Science 32, 335-348.

McMillen, D.P., 1995a. Spatial effects in limited dependent variable models. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Berlin: Springer-Verlag, pp. 189-228.

McMillen, D.P., 1995b. Selection bias in spatial econometric models. Journal of Regional Science 33(3), 417-438.

McMillen, D.P., Jarmin, R., Thorsnes, P., 1992. Selection bias and land development in the monocentric city model. Journal of Urban Economics 31, 273-284.

McMillen, D.P., McDonald, J.F., 1991. Urban land value functions with endogeneous zoning. Journal of Urban Economics 29, 14-27.

Mead, R., 1967. A mathematical model for the estimation of interplant competition. Biometrics 30, 295-307.

Metz, M.H., 1986. Different by Design: The Context and Character of Three Magnet Schools. Routledge Press, New York.

Metzler, L.L., 1996. Interdistrict Open Enrollment: A Statewide Study in Ohio. Doctoral Dissertation, Cleveland State University.

Moran, P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society B 10, 243-251.

Moran, P., 1950. A test for the serial independence of residuals. Biometrika 37, 178-181.

Moss, W.G., 1977. Large lot zoning, property taxes, and metropolitan area. Journal of Urban Economics 4(4), 408-427.

Naiman, R.J., Turner, M.G., 2000. A future perspective on North America's freshwater ecosystems. Ecological Applications 10, 958-970.

Navaro, P., Carson, R., 1991. Growth controls: policy analysis for the second generation. Policy Studies 24, 127-152.

Nelson, A.C, 1992. Characterizing exurbia. Journal of Planning Literature 6(4), 350-368.

Nelson, A.C., 1999. Comparing states with and without growth management. Analysis based on indicators with policy implications. Land Use Policy 16, 121-127.

Nelson, A.C., Moore, T., 1996. Assessing growth management policy implementation. Land Use Policy 13(4), 241-259.

Nelson, G.D., Hellerstein, D., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. American Journal of Agricultural Economics 79, 80-88.

Nelson, G.C., Harris, V., Stone, S.W., 2001. Deforestation, land use, and property rights: empirical evidence from Darien, Panama. Land Economics 77(2), 187-205.

Nelson, J.L., Carlson, K., Palonsky, B., 1993. Critical Issues in Education: A Dialectic Approach. New York: McGraw-Hill.

Newburn, D.A, Berck, P., Merenlender A., 2006. Habitat and open space at risk of land-use conversion: targeting strategies for land conservation. American Journal of Agricultural Economics 88(1), 28-42.

O'Brien, D.M., Murdoch, J.C., 2000. School choice in a large Texas school district. American Education Finance Association Annual Meeting, Austin, Texas.

Ohio Department of Education, 1991. A report on interdistrict and intradistrict open enrollment in Ohio. (Unpublished document available from the Ohio Department of Education).

Ohio Department of Education, 1993. Interdistrict open enrollment. (Unpublished document available from the Ohio Department of Education).

Ohio Department of Education, 2011. Open enrollment listing. Retrieved from http://education.ohio.gov

Ord, J., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120-126.

Ottensmann, J.R., 1977. Urban sprawl, land values, and the density of development. Land Economics 53(4), 389-400.

Pasha, H.A., 1996. Suburban minimum lot zoning and spatial equilibrium. Journal of Urban Economics 40(1), 1-12.

Pace, R.K., Barry, R., 1997. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291-297.

Peiser, R.B., 1989. Density and urban sprawl. Land Economics 65(3), 193-204.

Pinkse, J., Slade, M.E., 1998. Contracting in space: an application of spatial statistics to discrete-choice models. Journal of Econometrics 85, 125-154.

Plantiga, A.J., Irwin, E.G., 2006. Overview of empirical models. In: Bell, K.P., Boyle, K.J., Rubin, J. (Eds.), Economics of Rural Land-Use Change. Ashgate Studies in Environmental and Natural Resource Economics, pp. 113-134.

Poirier, D., Ruud, P.A., 1988. Probit with dependent observations. Review of Economic Studies 55, 593-614.

Porter, D., 1997. Managing Growth in America's Communities. Island Press, Washington, DC.

Quandt, R., 1956. A probabilistic model of consumer behavior. Quarterly Journal of Economic Behavior 70, 507-536.

Rapp, G.C., 2000. Agency and choice in education: does school choice enhance the work efforts of teachers? Education Economics 8(1), 37-63.

Reback, R., 2008. Demand (and supply) in an interdistrict public choice program. Economics of Education Review 27, 402-416.

Rincke, J., 2006. Policy innovation in local jurisdictions: testing for neighborhood influence in school choice policies. Public Choice 129, 189-200.

Ripley, B., 1981. Spatial Statistics. New York: Wiley.

Rothenberg, T., 1984. Approximate normality of generalized least squares estimates. Econometrica 52, 811-825.

Rouse, C.E., 1998. Private school vouchers and student achievement: an evaluation of the Milwaukee parental choice program. Quarterly Journal of Economics 113(2), 553-602.

Ruggles, D., 1997. Ohio School District Funding Gains and Losses as a Result of Interdisctrict Open Enrollment. Youngstown, OH: Youngstown State University.

Rusk, D., 1999. Inside Game/Outside Game. Washington DC: Brookings Institute Press.

Ruud, P.A., 1991. Extensions of estimation methods using the EM algorithm. Journal of Econometrics 49, 305-341.

Saavedra, L.A., 2000. A model of welfare competition with evidence from AFDC. Journal of Urban Economics 47, 248-279.

Saiger, A.J., 2010. The school district boundary problem. Urban Lawyer 42, 495-547.

Sargan, J., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26, 393-415.

Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. American Economic Review 80(3), 465-479.

Schatzki, T., 2003. Options, uncertainty, and sunk costs: an empirical analysis of land use change. Journal of Environmental Economics and Management 46, 86-105.

Schneider, B, Schiller, K.S., Coleman, J.S., 1996. Public school choice: some evidence from the national education longitudinal study of 1988. Educational Evaluation and Policy Analysis 18(1), 19-29.

Smith, A.G., 1995. Public school choice and open enrollment: implications for education, desegregation, and equity. Journal of Law and Education 24, 147-194.

Smith, K.B., Meier, K.J., 1995. The Case Against School Choice. Armonk, NY: M.E. Sharpe.

Staal, S., Baltenweck, I., Waithaka, M., deWolff, T., 2002. Location and uptake: integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. Agricultural Economics 27, 295–315.

Thurstone, L.L., 1927. A law of comparative judgment. Psychological Review 34, 273-286.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234–40.

Towe, C., Nickerson, C., Bockstael, N.E., 2008. An empirical examination of the timing of land conversions in the presence of farmland preservation programs. American Journal of Agricultural Economics 90(3), 613-626.

Train, K.E., 2007. Discrete Choice Models with Simulation. New York: Cambridge University Press.

Train, K., Ben-Akiva, M., Atherton, T., 1989. Consumption patterns and self-selecting tariffs. Review of Economics and Statistics LXIX, 383-391.

Upton, G., Fingleton, B., 1985. Spatial Analysis by Example. New York: Wiley.

Vijverberg, W.P., 1997. Monte Carlo evaluation of multivariate normal probabilities. Journal of Econometrics 76, 281-307.

Wallace, N.E., 1988. The market effects of zoning undeveloped land: does zoning follow the market? Journal of Urban Economics 23, 307-326.

Wang, X., Kockelman, K., 2009. Application of the dynamic spatial ordered probit model: patterns of land development change in Austin, Texas. Papers in Regional Science 88(2), 345-365.

White, H., 1984. Asymptotic Theory for Econometricians. New York: Academic Press.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434-449.

Williams, H., 1977. On the formation of travel demand models and economics evaluation measures of user benefit. Environmental Planning A 9, 285-344.

Witte, J.F., 1993. The Milwaukee private-school parental choice program. In: Rasell, E., Rothstein, R. (Eds.), School Choice: Examining the Evidence. Economic Policy Institute, Washington, D.C., pp. 69-110.

Witte, J.F., Sterr, T.D., Thorn, C.A., 1995. Fifth-year report: Milwaukee parental choice program. University of Wisconsin, Madison, WI.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA, MIT Press.

Young, T.W., Clinchy, E., 1992. Choice in Public Education. New York: Teachers College, Columbia University.

Ysseldyke, J.E., Lange, C.M, Delaney, T.J., 1992. School choice programs in the fifty states. Research report no.7. Enrollment options for students with disabilities. University of Minnesota. Retrieved from http://www.eric.ed.gov/PDFS/ED367104.pdf

Zhou, B., Kockelman, K., 2008. Neighborhood impacts on land use change: a multinomial logit model of spatial relationships. The Annals of Regional Science 42(2), 321-340.

BIOGRAPHICAL SKETCH

Ledia Guci was born in Korca, Albania. She received an Associate of Arts degree in agribusiness management from Dimitris Perrotis College of Agricultural Studies, Thessaloniki, Greece, and a Bachelor of Science degree in agricultural business from University of Arkansas in 2005. She joined the Food and Resource Economics Department at the University of Florida in the fall of 2006 to pursue graduate studies in agricultural economics with a concentration in international trade. After completing the coursework for the first year of the master's program and while working as a research assistant for the Florida Department of Citrus, she was offered funding to start the doctoral program in the same department. She was awarded a Master of Science degree in 2008 and a Doctor of Philosophy degree in August of 2011. Her research interests are in the areas of applied econometrics, spatial econometrics, natural resource and environmental economics, and urban and regional economics.