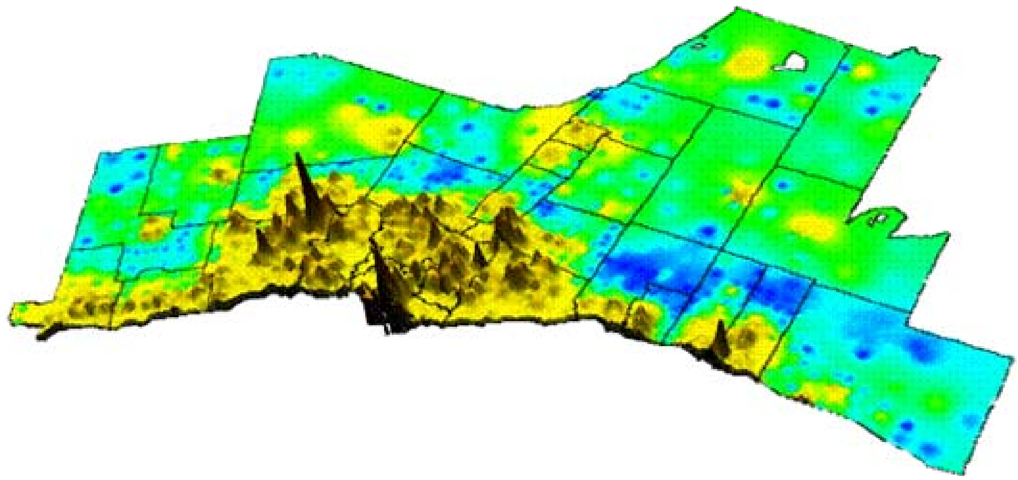


Applications of Spatial Multinomial Logit Model to Transportation Planning

Abolfazl Mohammadian, University of Illinois at Chicago
Pavlos S. Kanaroglou, McMaster University

Conference paper
Session XXX

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research

Lucerne, 10-15. August 2003

Applications of Spatial Multinomial Logit Model to Transportation Planning

Abolfazl Mohammadian
Department of Civil and Materials Engineering
University of Illinois at Chicago
Chicago, Illinois, USA

Phone: +1 312 413-2645
Fax: +1 312 996-2426
eMail: a.mohammadian@utoronto.ca

Pavlos S. Kanaroglou
School of Geography and Geology
McMaster University
Hamilton, Ontario, Canada

Phone: +1 905 525-9140 Ext. 23525
Fax: +1 905 546-0463
eMail: pavlos@mcmaster.ca

Abstract

This paper presents the process of derivation and development of a spatial multinomial logit model and its application to a housing type choice problem. Over the past few years, a relatively small body of research was developed that tries to capture the spatial and temporal dependencies across decision-makers and alternatives. While temporal dependencies are often considered especially in dynamic models, there has been relatively little work in the literature on incorporating spatial dependencies into qualitative dependent variables and discrete choice models, leading to inconsistent estimates. Part of the reason is that space is in general more complex to deal with than time. The basic idea presented in this paper is that decision-makers may influence each other, resulting in correlated choice behaviour over space. In this paper, spatial dependency terms are implemented in a standard multinomial logit framework. The results show that the spatial terms are statistically significant in the model and improve model fit.

Keywords

Spatial Logit Model, Spatial Dependency, Housing Choice Model, International Conference on Travel Behaviour Research, IATBR

Preferred citation

Mohammadian, Abolfazl, and Pavlos Kanaroglou (2003) Applications of Spatial Multinomial Logit Model to Transportation Planning, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. Introduction

Discrete choice models are widely used in economic, marketing, transportation and other fields to represent the choice of one among a set of mutually exclusive alternatives. Most discrete choice models are based on random utility maximization hypothesis. The development of discrete choice models represents a significant advance in the analysis of individual choice behaviour. Multinomial logit model is the most popular form of discrete choice model in practical applications. It is based on several simplifying assumption such as independently and identically Gumbel distribution (IID) of random components of the utilities and the absence of heteroscedasticity and autocorrelation in the model. Some of these assumptions do not appear unrealistic, but taken together it is often unrealistic to imagine any real world situations where all the conditions are satisfied. It has been shown that these simplifying assumptions limit the ability of the model to represent the true structure of the choice process. Recent research works contribute to the development of closed form models which relax the assumption of the multinomial logit model to provide a more realistic representation of choice probabilities within a closed form framework. Mixed logit and Generalized Extreme Value (GEV) models are examples of these alternative model structures (see Baht, 2002 for a detailed discussion).

Activity-based travel demand analysis postulates that “transportation is derived demand”. People travel to participate in different activities, which are distributed over space and time. Recent studies in travel behaviour research focus on activity location and spatial interaction of activities. These spatial interactions and dependencies (spatial autocorrelation) warrant modelling techniques that explicitly account for space. Spatial autocorrelation is defined as the dependency found in a set of cross-sectional observations over space. It occurs when individuals in population are related through their spatial location (Anselin, 1988). It may also occur when the choice decisions of individuals located in close proximity in space tend to be similar. If the choice set consists of spatial units, then alternatives closer to each other tend to be viewed by decision-makers as more similar than alternatives that are far apart.

Over the past few years, a relatively small body of research has developed that addresses the issue of spatial and temporal dependencies across decision-makers and alternatives. While researchers often account for temporal autocorrelation, spatial autocorrelation is generally ignored in discrete choice modelling, leading to inconsistent estimates (McMillen, 1995). Part of the reason is that research for the latter has not progressed to the point that the appropriate tools be readily available. Additionally, space is in general more complex to deal with than time. While time is one-dimensional and moves in one direction from past to present and to the future, space, in its simplest conceptualization, is two-dimensional and spatial processes can operate in any direction.

In this context, one might argue that decision-makers may influence each other, resulting in correlated choice behaviour over space. An individual selects one alternative from the available choice set on the basis of, among other things, the knowledge acquired from interactions with other decision-makers, such as colleagues, friends, or neighbours. This spatial dependency can be studied in several choice contexts including activity scheduling and land development choices. In activity scheduling, one may assume that the location of an individual influences his or her behaviour. Individuals face choice sets from which they select one alternative on the basis of knowledge they acquire through interactions with other decision-makers. It can be postulated that proximity to other decision-makers in space influences the decision process and in fact this influence increases with proximity.

Spatial autocorrelation can occur in many contexts. Several studies have estimated models of spatial dependence with continuous random variables, in the context of regression analysis. Anselin (1988) provides complete presentation of different approaches in this context including spatial weight matrices. Case (1992) developed a spatial model within a technology change context to show that the decision-making process of a farmer to adopt a new technology can be influenced by the expected profit of neighbours with whom there is a contact. Dubin (1992) used kriging or best linear unbiased prediction (BLUP) technique to predict housing prices in the presence of spatial autocorrelation. Dubin (1998) modelled the correlation structure itself, rather than the underlying process, and compared the resulting correlation structure with the one from the weight matrix technique.

There has been relatively little work in the literature on incorporating spatial dependencies into qualitative dependent variables and discrete choice models. One of the earliest attempts in this context is the work by Boots and Kanaroglou (1988), which incorporates the effect of spatial structure in discrete choice models of migration. Dubin (1995) implements this idea within a binary logit model applied to the diffusion of a technological innovation. In her model, the probability of adoption of new technology varies with the firm’s own characteristics and its interactions with previous adopters. She models the number of interactions as a diminishing function of geographic distance between firms. Paez and Suzuki (2001) used the binary logit model developed by Dubin (1995) and applied it to a land use problem and studied the effects of transportation on land use change.

McMillen (1995) investigated spatial effects in probit models. He found that heteroscedasticity causes inconsistent estimates in standard probit models and developed a heteroscedastic probit model using a Monte Carlo approach to capture the spatial effects. In a similar study, Berton and Vijverberg (1999) developed a conventional probit model on binary data where spatial dependencies (spatial lag or spatial error) were present.

In this paper, spatial dependency terms are implemented in a multinomial logit model. The model is applied to a housing type choice problem. It has been argued that the location of a housing project influences the type (e.g., detached vs. apartment) of new housing to be built in neighbouring locations. Additionally, it has been shown that the existing housing stock, as well as the location factors can affect the future housing developments in the same neighbourhood. This implies that the unobserved attributes of the neighbourhood (error terms) tend to be correlated.

This paper is structured in six sections. Section 2 briefly explains the model derivation and the approach employed in this study. Section 3 describes the data set used. Section 4 presents the process of development of spatial multinomial logit model and estimation results. Section 5 describes the analysis of the results obtained from maximum likelihood estimation of spatial multinomial logit model and comparison with standard multinomial logit model. Finally, Section 6 will present the conclusion and discussion.

2. Model Derivation

Discrete choice models are based on random utility theory, which assumes that the decision-maker's preference for an alternative is captured by the value of an index, called utility. A decision-maker selects the alternative from the choice set that has the highest utility value. Random utility models assume that decision-makers have perfect discriminating capability. However, the analyst will have limited information about an individual's utility level. The uncertainty introduced in this way must be taken into account (Ben-Akiva and Bierlaire, 1999). Equation 1 represents the utility of alternative i in the choice set C_n for decision-maker n (U_{in}), which is considered to be a random variable.

$$U_{in}=V_{in}+\varepsilon_{in} \quad [1]$$

It consists of an observed deterministic (or systematic) component of utility (V_{in}) and a randomly distributed unobserved component (ε_{in}) capturing the uncertainty. It is assumed that the alternative with highest utility is chosen. In order to account for spatial dependency, it is assumed that the systematic component of utility function (V_{in}) consists of two parts; the first part is a linear in the parameters function that captures the observed attributes of decision-makers n and alternatives

i, while the second term captures spatial dependencies across decision-makers. Utility of alternative i for decision-maker n is given as:

$$U_{in} = V_{in} + \varepsilon_{in} = (\sum \beta_i X_{in} + \sum_{s=1}^{S} \rho_{nsi} y_{si}) + \varepsilon_{in} \qquad [2]$$

where parameters β_i make up a vector of parameters (to be estimated) corresponding to X_{in}, the vector of observed characteristics of alternative i and decision-maker n. Parameters ρ make up a matrix of coefficients representing the influence that the choice of decision-maker s has on decision-maker n while choosing alternative i. S is the number of decision-makers who have influence on n. y_{si} will be set equal to unity if the decision-maker s has chosen alternative i, and zero otherwise. ρ can be modelled similar to an impedance function. In spatial statistics, it usually takes the form of a negative exponential function of the distance separating the two decision-makers (D_{ns}).

$$\rho_{nsi} = \lambda \exp(-\frac{D_{ns}}{\gamma}) \qquad [3]$$

where λ and γ are parameters to be estimated. The total influence that the choices of all other decision-makers have on decision-maker n can be modelled as:

$$Z_{in} = \sum_{s=1}^{S} \rho_{nsi} y_{si} \qquad [4]$$

Probability that decision-maker n would choose alternative i, rather than any other alternative j in the choice set, can be expressed as the probability that the utility of i is higher than that of any other alternative, conditional on knowing the systematic utility V_{jn} for all j alternatives in the choice set. Derivation of choice model proceeds in a fashion similar to that of the multinomial logit model (Ben-Akiva and Lerman, 1985).

$$P_{in} = \frac{\exp(\mu V_{in})}{\sum_{j \in C_n} \exp(\mu V_{jn})} \quad j \in C_n \qquad [5]$$

Appendix A provides the process of derivation of spatial multinomial logit model in details. The systematic utility function of alternative i for decision-maker n is given as:

$$V_{jn} = \sum_k \beta_k X_{kjn} + Z_{in} = \sum_k \beta_k X_{kjn} + \sum_{s=1}^{S} \rho_{\sin} y_{si} \qquad [6]$$

The log-likelihood function for a sample of size N is given by:

$$L^*(\beta) = ln\ (L(\beta)) = \sum_{n=1}^{N}\sum_{i\in C_n} \ln P_{in}^{y_{in}} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[V_{in} - \ln(\sum_{j\in C_n} \exp(V_{jn}))] \qquad [7]$$

Where y_{in} is a dummy variable such that y_{in}= 1 if alternative i is chosen by decision-maker n, and y_{in}= 0 otherwise.

To calculate the spatial dependency term ρ in Equation 3, one needs estimates of the parameters λ and γ. The value of these two parameters can be estimated directly by maximizing the likelihood function in Equation 7. Alternatively, parameters λ and γ can be obtained via a search procedure over a range of numbers by trying out different values of the parameter γ while estimating the value of λ as a standard parameter in logit model using a standard software, such as LIMDEP (Greene, 2002), and selecting the one that best fits the data.

The unknown parameters in this study were obtained directly by maximum likelihood estimation of the spatial multinomial logit model with the help of a computer program that was developed using GAUSS programming language. For the purpose of demonstrating the viability of the developed model, it is applied to a dataset of housing type development choices by residential real-estate developers.

3. The Data

The data set used in this study was compiled from numerous sources. Haider (2003) presents a detailed descriptive analysis of the data set used for this study. The housing starts data in the Greater Toronto Area (GTA) were obtained from RealNet Canada Inc. The housing starts dataset consists of records of development projects including information on the type, location, size, developer, and price of the projects constructed during January 1997 to April 2001 in GTA.

Zonal level socio-economic characteristics of the GTA were obtained from the 1996 Transportation Tomorrow Survey (TTS) database. TTS 1996 is a telephone-based, travel survey of 5% of households within the GTA that was undertaken in the autumn of 1996 (DMG, 1997). The survey covers household socio-economic information along with all one-day trips made by household members 11 years of age or older for a randomly selected weekday.

Accessibility indices for various types of activities were later developed for each TTS zone. Instead of using straight-line or network distances as impedance factors, average estimated travel times from the GTA traffic assignment model were used for each zone. It is assumed that the accessibility indices will capture the relative accessibility advantage of one TTS zone over the other for various types of activities (e.g. work, shopping, etc).

Land-use information and land inventory data were compiled from data obtained from PMA Brethour Inc as well as Ministry of Municipal Affairs and Housing in Ontario. Contiguity matrix and distances between centroids of traffic zones were estimated from GTA 1996 traffic zone map obtained from the Joint Program in Transportation, University of Toronto. Numerous other measures of spatial attractiveness of a TTS zone were developed using GIS data from Statistics Canada.

The sample used in this study comprises 1384 housing projects for which all required explanatory variables were available. The explanatory variables used in this study as well as their sample means and standard deviations are presented in Table 1. A total of 546 housing projects or 39.5 percent of the sample are detached houses. Semi-detached houses account for 241 or 17.4 percent of developments. Apartment projects are 237 or 17.1 percent, and remaining 360 projects (about 26 percent) are other types of developments (townhouses, row houses, etc).

Table 1 Variables Used in the Model

Variable	Mean	Std. Dev.
Price: price of the housing unit ($\times 10^5$ Canadian Dollar)	2.158	0.518
Development Charge: municipal charge for the unit ($\times 10^3$ CAD)	10.161	5.009
Intersection Density: (street intersections/100) ÷ zonal area	1.848	2.691
School Accessibility: mean weighted school accessibility index	60.449	23.182
Employment Accessibility: mean weighted emp. accessibility index	77.773	34.889
Residential Area: area zoned as residential (km^2)	0.287	0.284
Built-up: binary variable (1 if built-up area, 0 otherwise)	0.611	0.488
Inventory: inventory of residential units	243.517	459.055
D_{ij}: distance between centroids of zone i and adjacent zone j (km)	1.698	0.824

4. Housing Type Development Choice Model

Each decision–maker in this study is assumed to hold a parcel of land and is about to start a housing project. Developers are faced with the decision of what type of residential units to build (i.e., detached, semi-detached, condo, or townhouse). It can be postulated that this decision is influenced, to some extend at least, by nearby housing development projects. In other words, the existing housing stock, as well as the location factors will affect the future housing developments in the same neighbourhood. This implies that the unobserved attributes of the neighbourhood tend to be correlated.

This section describes the process of modelling and empirical results of a spatial multinomial logit model that predicts housing start types chosen by land developers.

4.1 Choice Set Specification

The choice set from which land developers make their choices is defined by the available alternatives in the dataset. The dataset used for this study contains usable records of housing projects in GTA that started between 1997 and 2001. Initially, there were seven distinct housing types in the database. These were: single-detached house, semi-detached house, townhouse, row house, apartment, condo in high-rise, and others types of houses. Descriptive analysis revealed that some of these housing types are very uncommon. Therefore, seven housing types of the dataset were aggregated into four groups of detached, semi-detached, apartments, and others.

The dataset extracted to develop this model contains 1384 unweighted observations of housing start projects in land parcels. Developers face the decision to select housing type from four alternatives available in the choice set. It is assumed that all choices are available to all decision-makers.

4.2 Utility Function Specification

The unknown parameters of the spatial multinomial logit model were obtained directly by a maximum likelihood estimation of Equation 7. In order to specify the utility functions, it should be decided which variables to be included and in what form. Variables that can enter the utility functions include: choice attributes, decision-maker attributes, socioeconomic characteristics, and combinations of these variables. These variables can enter the utility functions in generic or alternative-specific form in an arbitrary number of alternatives. Intersection density, school and job accessibility indices, land-use related variables, and inventory of residential units in the zone are used as alternative specific variables.

Alternative specific constants, which capture the systematic impact of omitted variables in the utility function, were also included in the utility functions of the model. Price of the housing unit and development charge were two variables representing attributes of alternatives. The variable "development charge" is the municipal tax for different types of housing projects. Unfortunately no variable representing the attributes of decision-makers (developers) were available to be included in the model.

Additionally, a spatial dependence term, Z_{in}, is introduced to the utility function as shown in Equation 6. This term is a function of distances (D_{ij}) separating the housing project from adjacent projects of similar housing type (see Equation 4).

4.3 Estimation Results

The model has been estimated with and without the spatial dependency term using the same set of explanatory variables defined in Table 1. The results of maximum likelihood estimation of both multinomial logit (MNL) and spatial multinomial logit (SPNL) models are summarized in Table 2.

5. Analysis of the Results

Almost all parameters are statistically significant at 95% confidence level or better. Adding the spatial dependence term improves the overall goodness of fit of the model. The standard multinomial logit model has an adjusted log-likelihood ration (ρ^2) of 0.226 when comparing the log-likelihood at zero and log-likelihood at convergence. The constants alone contribute 0.045 of the 0.226, suggesting the attributes in the utility expressions play an important role in explaining the choice behaviour. This indicates a good model fit.

After adding the spatial dependency term to the model, the log-likelihood function value of -1478.77 increased to –1445.97 and the ρ^2 for the spatial logit model improved to 0.243. This presents the robustness of spatial logit model formulation and confirms the importance of the spatial dependency factors in explaining developer's housing type choice behaviour. This also indicates that information collected regarding housing start projects should include some level of spatial attributes.

Table 2 Estimation Results of Multinomial Logit and Spatial Multinomial Logit Models

		MNL		SMNL	
Variable	Alternative[1]	Parameter	t-statistic	Parameter	t-statistic
Price	D, S, O	0.148	1.615	0.160	1.641
Development Charge	D	-0.140	-2.215	-0.143	-2.268
	S	-0.167	-2.897	-0.162	-2.804
	O	-0.204	-3.257	-0.196	-3.119
	A	-0.288	-3.231	-0.276	-3.110
Intersection Density	D, S	-0.341	-5.359	-0.334	-5.202
School Accessibility	D, S	0.106	4.362	0.103	4.235
Employment Accessibility	D	-0.084	-4.418	-0.084	-4.378
	S	-0.074	-3.858	-0.074	-3.838
	A	0.049	8.195	0.043	7.142
Residential Area	S	-0.860	-2.863	-0.873	-2.898
	O	-1.085	-4.057	-1.087	-4.057
Built-up	A	0.470	3.067	0.424	2.760
Inventory	A	-0.005	-3.296	-0.004	-3.174
Alternative Specific Constant	D	4.572	5.821	4.209	5.391
	S	4.001	4.503	3.637	4.107
	O	4.731	6.216	4.216	5.539
λ	D, S, O, A			0.528	4.117
γ	D, S, O, A			1.953	3.528
Number of observations		1384		1384	
Log-likelihood at zero		-1918.63		-1918.63	
Log-likelihood constant-only model		-1832.11		-1832.11	
Log-likelihood at convergence		-1478.77		-1445.97	
Log-likelihood ratio (adjusted ρ^2)		0.226		0.243	

[1]Alternatives: Detached (D), Semi-Detached(S), Apartment (A), and Others (O)

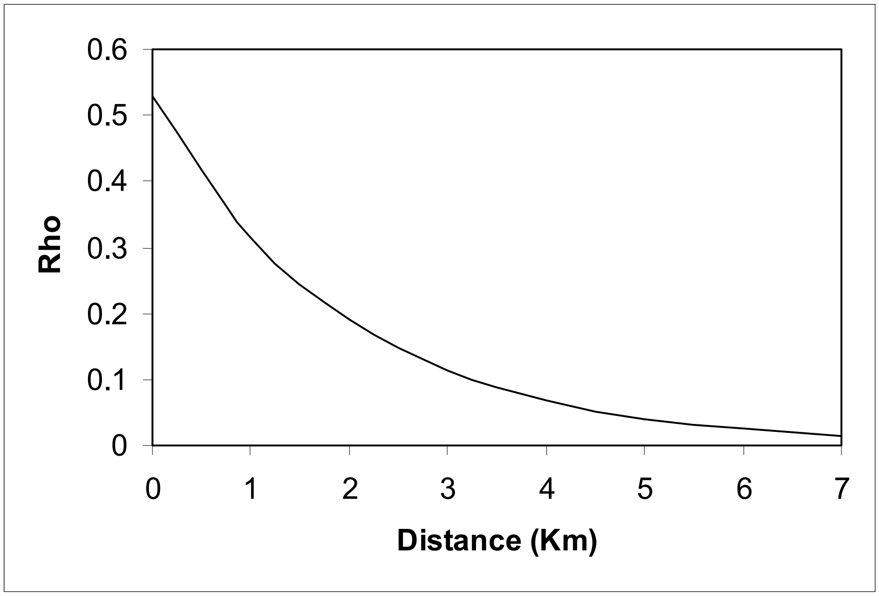
The signs of all utility parameters seem to be correct and unambiguous. We would expect that a negative sign would be associated with those attributes that may cause negative utility. Based on the evidence provided by the review of the literature, we expect the price of the housing unit to be associated with its development. In order to maximize the profit, developers tend to be interested to build housing units that are likely to be sold for a higher price.

Positive sign for the parameter of the price variable confirms this hypothesis. Additionally, it can be hypothesized that the higher development charges for a housing type, the lower the probability of building that particular type. As expected municipal development charge variable generated parameters with negative sign. This indicates that development charge is negatively associated with the choice since it enters the model as a 'cost' variable.

The model indicates that within the developed parts of the urban area, where street networks are highly developed and the intersection density is higher, the chance of building apartments is higher, while the chance of building detached and semi-detached houses is lower. The parameters for detached and semi-detached houses for the intersection density variable were almost identical, suggesting that we could save one degree of freedom by imposing an equality restriction on these two utility parameters, treating them as generic to these two alternatives. The same argument is true for the school accessibility variable, where higher accessibility to school will increase the chance of building detached and semi-detached houses. This is true since families with young children usually occupy these types of dwelling units. Higher accessibility to jobs and employment will increase the chance of building apartments, while it will decrease the chance of building detached and semi-detached houses. This is probably due to the fact that employment opportunities within urban area are usually closer to the CBD and the chance of building high-rise buildings and apartment units are higher in central areas. It has also been shown that if the total residential area is larger within the study zone, the chance of building semi-detached and other types of buildings will be lower. If the location of the project is within a built-up area, where the availability of undeveloped parcels of land is limited, the new project is more likely to be apartments. The parameter of the inventory of residential units in the utility function for apartments is negative in the model. Large number of housing units in a zone suggest that there are large tracts of developable land in that zone, making it more attractive to detached or semi-detached type developments and less attractive for apartment type developments.

The spatial dependency factor is a generic variable with positive signs for both parameters in all alternatives. It generated the expected signs and magnitudes of parameters. The positive sign of λ in the model indicates that the existence of similar housing type in adjacent zones is directly associated with the development type choice. In order to test the size and extent of the spatial effects, spatial parameters λ and γ resulted from this model are used to present a distance-decay curve as illustrated in Figure 1. The curve indicates that development projects within a 4-km buffer will have direct impact on the choice of the project type. Highly significant t-statistics for these two parameters suggest that neighbourhood effects are very important in the model developed in this study.

Figure 1 Distance-Decay curve



$$\rho_{nsi} = 0.528 \times \exp(-\frac{D_{ns}}{1.953})$$

6. Conclusions

This study presents the process of derivation and development of a spatial multinomial logit model and its application to a housing type choice problem of real-estate developers. Spatial dependency terms are employed in a standard multinomial logit framework to capture spatial interactions across projects. The results show that the spatial terms are statistically significant in the model. Additionally, the model captures interactions between development type choice behaviour and the existing land-use and transportation infrastructure.

Spatial dependency can be studied in several contexts. In addition to the land-use applications (with physical distance as impedance term) similar to the one developed here, it is possible to apply the model to other transportation and choice problems. For example, in activity scheduling, it can be argued that the location of an individual influences his or her behaviour. Every individual picks his or her available choice set and selects the best alternative based on the knowledge he or she acquires through interactions with other decision-makers (e.g. colleagues or friends) who are located at diverse points. The closer the other decision-maker, the higher his or her influence on individual's choice. The measure of closeness can be physical distance as well as non-physical measures (e.g. similarity indices).

In order to account for unobserved response heterogeneity, it is also possible to implement the spatial logit model within a mixed logit framework (see Ben-Akiva and Bolduc, 1996) and estimate the value of parameters as random terms in the model. This remains a task for future research on this topic.

7. Acknowledgement

The authors thank Prof. Murtaza Haider of the McGill University, for compiling and providing the dataset and for his intellectual and practical input that tangibly enhanced this study. In addition, authors are grateful to RealNet Canada, PMA Brethour, Ministry of Municipal Affairs and Housing, Urban Development Institute and other main providers of the data used in this study. The research reported in this paper is supported by a major collaborative research initiative grant from Social Science and Humanities Research Council of Canada (SSHRC) and the Canada Research Chair in Spatial Analysis held by Professor Kanaroglou.

8. References

Anselin, L. (1988) Spatial Econometrics: Methods and Models, Dordrecht: Kluwer Academic Publishers.

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

Ben-Akiva, M. and D. Bolduc (1996) Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure, 3rd International Choice Symposium, Columbia University.

Ben-Akiva, M. and M. Bierlaire (1999) Discrete Choice Methods and Their Applications to Short Term Travel Decisions, in R.W. Hall (Ed.) *Handbook of Transportation Science*, 5-33, Kluwer Academic Publishers, USA.

Berton, K.J., and W.P.M. Vijverberg (1999) Probit in a Spatial Context: A Monte Carlo Analysis, in L. Anselin, R. Florax and S. Rey (Eds.), *Advances in Spatial Econometrics, Methodology, Tools and Applications* (forthcoming), Springer Verlag, Berlin.

Bhat, C.R. (2002) Recent Methodological Advances Relevant to Activity and Travel Behavior, in H.S. Mahmassani (Ed.) *In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*, 381-414, Elsevier Science, Oxford, UK.

Boots, B.N., and P.S. Kanaroglou (1988) Incorporating the Effect of Spatial Structutre in Discrete Choices Models of Migration, Journal of Regional Science, 28, 495-507.

Case, A. (1992) Neighborhood Influence and Technological Change, Regional Science and Urban Economics, 22, 491-508.

DMG (1997) TTS Version 3: Data Guide, Research Report, Data Management Group, Joint Program in Transportation, University of Toronto.

Dubin, R.A. (1992) Spatial Autocorrelation and Neighborhood Quality, Regional Science and Urban Economics, 22, 433-452.

Dubin, R.A. (1995) Estimating Logit Models with Spatial Dependence, in L. Anselin and R. Florax (Eds.) *New Directions in Spatial Econometrics*, 229-242, Springer-Verlag, Heidelberg.

Dubin, Robin A. (1998) Spatial Autocorrelation: A Primer, in *Journal of Housing Economics*, 7, 304-327.

Greene W.H. (2002) NLOGIT Version 3.0 User's Manual, Econometric Software Inc., Plainview, NY, USA.

Haider, M. (2003) Spatio-temporal Modelling of Housing Starts in the Greater Toronto Area, Ph.D. Thesis, Dept. of Civil Eng., University of Toronto.

McMillen, D.P. (1995) Spatial Effects in Probit Models: A Monte Carlo Investigation, in L. Anselin and R. Florax (Eds.) *New Directions in Spatial Econometrics*, 189-228, Springer-Verlag, Heidelberg.

Paez, A., and J. Suzuki (2001) Transportation Impacts on Land Use Changes: An Assessment Considering Neighborhood Effects, Journal of Eastern Asia Society for Transportation Studies, 4, 47-59, October 2001.

Appendix A: Derivation of Spatial Multinomial Logit Model

Probability that decision-maker n would choose alternative *i,* rather than any other alternative *j* in the choice set, is given by the fact that the utility of *i* is higher than any other alternative, conditional on knowing the observed (or systematic) utility V_{jn} for all *j* alternatives in the choice set. This probability can be expressed as:

$$P_{in} = P\,[(V_{in} + \varepsilon_{in}) \geq \max_{j \in C_n}(V_{jn} + \varepsilon_{jn})\,] = P[\,(V^*_{jn} + \varepsilon^*_j) - (V_{in} + \varepsilon_{in}) \leq 0] \qquad \text{[A1]}$$

The multinomial logit model arises when the error terms in the utility function are assumed to be independently and identically distributed (IID) Type I (Gumbel) distributed with parameters (η_i, μ). Based on the properties of the Gumbel distribution max(ε_{1n}, ε_{2n},…, ε_{jn}) will be also Gumbel distributed with parameters ($\mu^{-1}\ln \Sigma_j \exp(\mu\eta_j)$, μ). If ε_1 and ε_2 are independent Gumbel distributed with parameters (η_1, μ), and (η_2, μ) respectively, then $\varepsilon^{**} = \varepsilon_1 - \varepsilon_2$ is logistically distributed:

$$F(\varepsilon_1 - \varepsilon_2) = F(\varepsilon^{**}) = \frac{1}{1 + \exp(\mu(\eta_2 - \eta_1 - \varepsilon^{**}))} \qquad \text{[A2]}$$

Thus, The probability that a given decision-maker *n* chooses alternative *i* within the choice set C_n is given by Equation A3.

$$P_{in} = \frac{1}{1 + \exp(-\mu(V_{in} - V^*_{jn}))} = \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \exp(\mu V^*_{jn})}$$

$$= \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \exp(\mu(\frac{1}{\mu}\ln \sum_{j \in C_n} \exp(\mu V_{jn})))} = \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \sum_{j \in C_n} \exp(\mu V_{jn})}$$

$$= \frac{\exp(\mu V_{in})}{\sum_{j \in C_n} \exp(\mu V_{jn})} \quad j \in C_n \qquad \text{[A3]}$$

and the systematic (observed) utility function in spatial multinomial logit model is:

$$V_{jn} = \sum_k \beta_k X_{kjn} + Z_{in} = \sum_k \beta_k X_{kjn} + \sum_{s=1}^{S} \rho_{\sin} y_{si} \qquad \text{[A4]}$$

The log-likelihood function for a sample of size N is given by:

$$L^{*}(\beta) = ln\ (L(\beta)) = \sum_{n=1}^{N}\sum_{i\in C_n}\ln P_{in}^{y_{in}} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[V_{in} - \ln(\sum_{j\in C_n}\exp(V_{jn}))] \qquad \text{[A5]}$$

Where y_{in} is a dummy variable such that y_{in}= 1 if alternative i is chosen by decision-maker n, and y_{in}= 0 otherwise. In order to obtain the maximum likelihood estimates of the parameters, Equation A5 should be maximized with respect to parameters β*'s*, λ, and γ. The usual approach is to use the Newton-Raphson technique that requires the first and second order derivatives of the log-likelihood function.

A 1: First and Second Order Derivatives

First order derivatives of $L^{*}(\beta)$ are given by:

$$\frac{\partial L^{*}(\beta)}{\partial \beta_{k}} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[X_{kin} - \frac{\sum_{j\in C_n}(e^{V_{jn}}X_{kjn})}{\sum_{j\in C_n} e^{V_{jn}}}] = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[X_{kin} - \frac{\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}} X_{kjn})}{\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}}}] \qquad \text{[A6]}$$

$$\frac{\partial L^{*}(\beta)}{\partial \lambda} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[\sum_{s=1}^{S}(\frac{\partial \rho_{\sin}}{\partial \lambda} y_{si}) - \frac{\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}} . \sum_{s=1}^{S}(\frac{\partial \rho_{\sin}}{\partial \lambda} y_{si}))}{\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}}}] \qquad \text{[A7]}$$

$$\frac{\partial L^{*}(\beta)}{\partial \gamma} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}[\sum_{s=1}^{S}(\frac{\partial \rho_{\sin}}{\partial \gamma} y_{si}) - \frac{\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}} . \sum_{s=1}^{S}(\frac{\partial \rho_{\sin}}{\partial \gamma} y_{si}))}{\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn} + Z_{jn}}}] \qquad \text{[A8]}$$

and second order derivatives of $L^{*}(\beta)$ are given by:

$$\frac{\partial^2 \mathrm{L}^*(\beta)}{\partial\beta_\mathrm{k}\partial\beta_l} = -\sum_{n=1}^{N}\sum_{i\in C_n} y_{in}\Bigg[\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn}X_{ljn}))(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}$$

$$-\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{ljn}))(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}\Bigg] \qquad \text{[A9]}$$

$$\frac{\partial^2 \mathrm{L}^*(\beta)}{\partial\beta_\mathrm{k}\partial\gamma} = -\sum_{n=1}^{N}\sum_{i\in C_n} y_{in}\Bigg[\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn}\sum_{s=1}^{S}(\frac{\partial\rho_{\mathrm{sin}}}{\partial\gamma} y_{si})))(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}$$

$$-\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial\rho_{\mathrm{sin}}}{\partial\gamma} y_{si})))(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}\Bigg] \qquad \text{[A10]}$$

$$\frac{\partial^2 \mathrm{L}^*(\beta)}{\partial\beta_\mathrm{k}\partial\lambda} = -\sum_{n=1}^{N}\sum_{i\in C_n} y_{in}\Bigg[\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn}\sum_{s=1}^{S}(\frac{\partial\rho_{\mathrm{sin}}}{\partial\lambda} y_{si})))(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}$$

$$-\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial\rho_{\mathrm{sin}}}{\partial\lambda} y_{si})))(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}} X_{kjn})}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}\Bigg] \qquad \text{[A11]}$$

$$\frac{\partial^2 \mathrm{L}^*(\beta)}{\partial\lambda\partial\gamma} = \sum_{n=1}^{N}\sum_{i\in C_n} y_{in}\{\sum_{s=1}^{S}(\frac{\partial^2\rho_{sjn}}{\partial\lambda\partial\gamma}y_{sj}) -$$

$$[\frac{(\sum_{j\in C_n}[e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial\rho_{sjn}}{\partial\gamma}y_{sj})\sum_{s=1}^{S}(\frac{\partial\rho_{sjn}}{\partial\lambda}y_{sj}) + e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial^2\rho_{sjn}}{\partial\lambda\partial\gamma}y_{sj})])}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2} \times (\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})]$$

$$+\frac{(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial\rho_{sjn}}{\partial\gamma}y_{sj})))(\sum_{j\in C_n}(e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}}\sum_{s=1}^{S}(\frac{\partial\rho_{sjn}}{\partial\lambda}y_{sj})))}{(\sum_{j\in C_n} e^{\sum_{k=1}^{K}\beta_k X_{kjn}+Z_{jn}})^2}\} \qquad [A12]$$