Homebuilders Choice Behaviour Analysis[*]

Abolfazl Mohammadian
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, Illinois 60607-7023

Murtaza Haider   Pavlos S. Kanaroglou
Ted Rogers School of   School of Geography and Geology
Retail Management   McMaster University
Ryerson University   1280 Main Street West
350 Victoria Street   Hamilton, ON, L8S 4K1
Toronto, ON, M5B 2K3

The housing sector in Canada is an important and significant component of the economy. Investment in new housing construction in Canada totaled C$79.8 billion in 2006, representing almost 6.8% of Canadian GDP. In 2006, construction began on 227,395 new housing units in Canada. With 1.6 million new dwellings built each year, the impact of new housing construction is even more pronounced in the United States. Whereas billions of dollars are invested in new housing construction every year, it is rather surprising that the development behaviour of homebuilders has not been subjected to the same level of empirical scrutiny that has been given to housing demand.

We argue that in addition to the role of highway construction and improvement projects, urban form, to a large extent, is determined by explicit choices made by homebuilders. The intensity of development, measured as housing units per acre, and type of new housing to be developed, such as single-

family detached or apartments, are the choices made by homebuilders, which end up defining the urban landscape. Moreover, homebuilders also determine when to build and how much of new housing to build. While the abovementioned decisions made by homebuilders are in response to certain market signals they receive from the homebuyers' consumption behaviour, the homebuilders, however, operate with a certain degree of freedom about the choices they make. Whereas households decide where to live and what type of housing to occupy, homebuilders define the choice set that is made available to households. Thus, to a certain extent, the household's housing choices are conditional upon the choices already made by homebuilders. It is therefore very surprising that the type, location, and timing of new housing construction, which are decisions made by homebuilders, have been the focus of only a handful of studies.

The heterogeneity in the size and type of homebuilding firms suggest that there is no one standard type of homebuilder. In the Greater Toronto Area, the large homebuilding firms build in excess of 7,000 units per year, whereas the small homebuilders may build fewer than 10 units per year. Despite the variance in type and size of homebuilding firms, the homebuilders represent certain common traits. For instance, the greenfield low-density housing developments on the urban fringe, which resulted in suburbia, suggest some behavioural conformity among homebuilders. Rows of cookie-cutter housing built by different homebuilders suggest that the type of new housing development is influenced by other residential development projects that have been built in the recent past. Again, these spatial dependencies impacting homebuilders' choices have not been accounted for in the empirical studies.

Land-use and housing choice models rely on behaviours that vary over space and time resulting in spatial and temporal dependencies across decision-makers and alternatives. While researchers often account for temporal autocorrelation, the spatial dimension is frequently ignored in discrete choice models, leading to inconsistent estimates (McMillen 1995). One possible explanation might be that space is in general more complex to deal with than time. While time can be considered as one-dimensional, even in its basic form, space is two-dimensional. If spatial dependencies exist, it can be argued that decision-makers may influence each other, resulting in correlated choice behaviour over space. We specifically argue that proximity to other decision-makers (homebuilders in this application) in space influences the decision process and that this influence increases with proximity.

Earlier attempts to account for spatial dependence focused largely on continuous dependant variables, in the context of regression analysis (Anselin 1988; Case 1992; Dubin 1992, 1998). There has been relatively little work in the literature on incorporating spatial dependencies into discrete choice models. Boots and Kanaroglou (1988) incorporated the effect of spatial structure in discrete choice models of migration. Few other studies investigated spatial effects in conventional binary probit models where spatial dependencies were present (McMillen 1995; Berton and Vijverberg 1999). Dubin (1995) developed a spatial binary logit model to predict the diffusion of a technological innovation. In her model, the probability of adoption of a new technology varies with the firm's own

characteristics and its interactions with previous adopters. Paez and Suzuki (2001) tested the application of this spatial binary logit model to a land use problem, considering the effects of transportation on land use changes. Mohammadian and Kanaroglou (2003) expanded the binary choice model into a more general form, deriving a spatial multinomial logit model (SMNL) and tested it on a housing type choice problem. Bhat and Guo (2003) developed a mixed spatially correlated logit model for residential location choices. Their model combines the generalized extreme value (GEV) model and a mixed multinomial logit model to allow spatial correlation across alternatives.

In this paper, spatial dependency terms are implemented in a mixed (random parameter) logit framework to form a spatial mixed logit model (SML), allowing utility coefficients and spatial dependency term parameters to differ across decision makers (homebuilders). The model is applied to a housing type choice problem for new housing projects. The results of the model present a substantial improvement, in terms of model fit, over spatial logit and standard multinomial logit models.

This paper is structured in six sections. The second section briefly explains the model specification. The third section describes the data set. The fourth section presents the process of model development and estimation results. The fifth section describes the analysis of the results and, the sixth and final section presents the conclusion and discussion.

Model Specification

Random utility based discrete choice models have made their way into many disciplines including transportation, marketing, and other fields. Multinomial logit model (MNL), the most popular form of discrete choice models in practical applications, is based on several simplifying assumptions. These include the independent and identical Gumbel distribution (IID) of random components of the utilities and the absence of heteroscedasticity and autocorrelation in the model. It has been shown that these simplifying assumptions limit the ability of the model to represent the true structure of the choice process. Recent papers have contributed to the development of closed form models that relax some of these assumptions and provide a more realistic representation of choice probabilities. Mixed logit (ML) and Generalized Extreme Value (GEV) models are examples of these alternative structures (see Bhat (2002) for a detailed discussion).

Discrete choice models assume that decision-maker's preference for an alternative is captured by the value of an index, called utility and a decision-maker selects the alternative from the choice set that has the highest utility value (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 (Ben-Akiva and Lerman 1985).

$$U_{in} = V_{in} + \varepsilon_{in} \qquad\qquad (1)$$

The utility function consists of an observed deterministic (or systematic) component of utility (V_{in}) and a randomly distributed unobserved component (ε_{in}) that captures the uncertainty. It is assumed that the alternative with the highest utility is chosen.

In some applications of discrete choice models, such as housing choice, it is quite likely that choices are correlated in space. Spatial correlation is defined as the dependency found in a set of observations over space (Anselin 1988). In the context of spatial choice, it occurs when individual decision-makers are related through their spatial proximity. It may also occur when the choice set consists of spatial entities in varying levels of proximity with each other. A common approach in the spatial analysis literature for capturing the correlations across alternatives is to allow alternatives that are contiguous to be correlated. The unobserved spatial correlation across alternatives can be accounted for by utilizing a mixed GEV model as shown by Bhat and Guo (2003). On the other hand, to account for spatial dependency across decision makers, the utility function defined in Equation 1 can be modified to consider spatial interactions and dependencies.

It can be assumed that the systematic component of utility function (V_{in}) consists of two parts; the first part is a linear-in-parameters function that captures the observed attributes of decision-maker n and alternative i, while the second term captures spatial dependencies across decision-makers. Utility of alternative i for decision-maker n is given as:

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

where parameters β_i make up a vector of parameters (to be estimated) corresponding to X_{in}, the vector of the observed characteristics of alternative i and decision-maker n. Parameter ρ makes 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 some influence on n. y_{si} will be set to unity if the decision-maker s has chosen alternative i, and zero otherwise. ρ can be modeled 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}).

$$\rho_{nsi} = \lambda \, \exp\left(-\frac{D_{ns}}{\gamma} \right) \qquad\qquad (3)$$

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

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

Derivation of choice model proceeds in a fashion similar to that of the multinomial logit model.

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

Mohammadian and Kanaroglou (2003) provide the process of derivation of the spatial multinomial logit (SMNL) model in detail. A summary of the approach is presented here. The systematic utility function of alternative i for decision-maker n is given as:

$$V_{jn} = \sum_{k} \beta_k X_{kjn} + Z_{in} = \sum_{k} \beta_k X_{kjn} + \sum_{s=1}^{S} \rho_{sin} y_{si} \tag{6}$$

To estimate the spatial dependency term ρ in Equation 3, one needs to estimate the parameters λ and γ. These parameters, along with the vector of parameters β, can be estimated directly through maximum likelihood estimation.

This simple spatial multinomial logit model can be expanded into a mixed logit (ML) framework to form a spatial mixed logit model (SML) that accounts for heterogeneity across decision-makers while allowing spatial correlations across contiguous decision makers. The Mixed Logit model has been introduced by Ben-Akiva and Bolduc (1996) to bridge the gap between logit and probit models by combining the advantages of both techniques. A small yet growing number of empirical studies make use of the ML model. The earlier studies include Revelt and Train (1998), Bhat (1997 and 2000), and Brownstone et al (2000). In order to illustrate this type of model and to derive a SML, we need to modify equation 2:

$$U_{int} = \alpha_{in} + \gamma_i W_n + \beta_{in} X_{int} + \sum_{s=1}^{S} \rho_{sin} y_{si} + \varepsilon_{int} \tag{7}$$

where α_{in} is a constant term and captures an intrinsic preference of decision-maker n for alternative i, γ_i, W_n captures systematic preference heterogeneity as a function of socio-demographic characteristics, and X_{int} is the vector of attributes describing alternative i for decision-maker n, in the choice situation t. The vector of coefficients β_{in} is assumed to vary in the population, with probability density given by $f(\beta \mid \theta)$, where θ is a vector of the true parameters of the taste distribution. A spatial dependency term is also introduced to the equation. If the ε's are IID type I extreme value, the probability that decision-maker n chooses alternative i in a

choice situation *t* is given by:

$$P_{nt}(i|\beta_n) = \frac{\exp(\alpha_{in} + \gamma_i W_n + \beta_{in} X_{int} + \sum_{s=1}^{s} \rho_{sin} y_{si} + \varepsilon_{int})}{\sum_{j \in C_{nt}} \exp(\alpha_{jn} + \gamma_j W_n + \beta_{in} X_{jnt} + \sum_{s=1}^{s} \rho_{sin} y_{si} + \varepsilon_{jnt})} \tag{8}$$

Note that the probability in the above equation is conditional on the distribution of β_{in}. As mentioned earlier, a subset of all of α_{in} alternative-specific constants and the vector of parameters β_{in} can be randomly distributed across decision-makers. An important element of these random parameter models is the assumption regarding the distribution of each of the random coefficients. It may seem natural to assume a normal distribution. However, one coefficient might then be negative for some individuals and positive for others. For most of the variables, it is reasonable to expect that all respondents have the same sign for their coefficients, for example the coefficient for the cost variable should always be non-positive. For this type of coefficient, a more reasonable assumption would be to assume a log-normal distribution. For a more detailed treatment of preference heterogeneity, see Bhat (2000).

Since actual tastes are not observed, the probability of observing a certain choice is determined as an integral of the appropriate probability formula over all possible values of β_n weighted by its density. Therefore, the unconditional probability of choosing alternative *i* for a randomly selected decision-maker *n* is then the integral of the conditional multinomial choice probability over all possible values of β_n,

$$P_{nt}(i|\theta) = \int_{\beta} P_{nt}(i|\beta_n) f(\beta|\theta) d\beta \tag{9}$$

In this simple form, the utility coefficients vary over decision-makers, but are constant over the choice situations for each decision-maker. In general, the integral cannot be analytically calculated and must be simulated for estimation purposes. In order to develop the likelihood function for parameter estimation, we need the probability of each sample individual's sequence of observed choices. If T_n denotes the number of choice occasions observed for decision-maker *n*, the likelihood function for decision-maker *n*'s observed sequence of choices, conditional on β, is:

$$L_n(\beta) = \prod_{t-1}^{T_n} \prod_{i-1}^{I} [P_{nt}(i|\beta)]^{y_{nit}} \tag{10}$$

where y_{nit} will take the value of 1 if the nth decision-maker chooses alternative *i* on choice occasion *t* and zero otherwise. The unconditional likelihood function of the

choice sequence is:

$$L_n(\theta) = \int_{\beta} L_n(\beta) f(\beta|\theta) d\theta \tag{11}$$

The goal of the maximum likelihood procedure is to estimate θ. The log-likelihood function is:

$$L(\theta) = \sum_n \ln Ln(\theta) \tag{12}$$

Exact maximum likelihood estimation is not available and simulated maximum likelihood is to be used instead. In this method, all parameters are estimated by drawing pseudo-random realizations from the underlying error process. The individual likelihood function is then approximated by averaging over the different $L_n(\beta)$ values to estimate a simulated likelihood function. The parameter vector θ is estimated as the vector value that maximizes the simulated function. For detailed discussion of this method see Louviere et al (2000) and Bhat (2000).

The unknown parameters in this study were obtained directly by maximizing the simulated likelihood function of the SML model.

Data

The data on the type of new housing construction and its determinants were compiled from numerous sources. Haider (2003, 2004) presents a detailed descriptive analysis of the data set used for this study. The housing construction dataset consists of records of new housing developments including information on the type, location, size, and price of new housing constructed during January 1997 and April 2001 in the Greater Toronto Area (GTA). The database included all new housing developments that included a minimum of ten new housing units.

Zonal level socio-economic characteristics 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 (Data Management Group 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 capture the relative accessibility advantage of one TTS zone over the other for various types of activities (e.g. work and shopping).

Contiguity matrix and distances between centroids of traffic zones were calculated from the GTA 1996 traffic zone map obtained from the Joint Program in Transportation, University of Toronto. Numerous other measures of spatial attractiveness of zones were developed using GIS data from Statistics Canada.

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
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

It is worth noting that, for simplicity, it has been assumed that the accessibility measures and other parameters of the estimation data will not change over the study period (1997 to 2001). While this assumption might be a point of concern, it is necessary due to the limitations of available data.

The final sample used in this study comprised 1384 new housing projects for which all required explanatory variables were available. Each project represented the homebuilders' decision to build a particular type of housing units. 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 % of the sample are single-family detached (SFD) houses. Semi-detached (SD) houses account for 241 or 17.4 % of developments. Apartment projects are 237 or 17.1 %, and the remaining 360 projects (about 26 %) are other low-rise developments (e.g. townhouses and row houses). The 1384 housing projects account for the construction of 113,000 new housing units.

Model Development

Homebuilders are faced with the decision of what type of residential units to build (i.e., SFD, SD, condominium, 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 may be correlated.

The homebuilders' choice set is defined by the available alternatives in the dataset. Initially, there were seven distinct housing types in the database. These were: single-family SFD, SD, townhouse, row house, apartment, condo in high-rise, and other types of housing units. 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 SFD, SD, apartments, and other low-rise housing. The dataset extracted to develop this model contains 1384 unweighted observations of new housing projects. Developers face the decision to select housing type from four alternatives available in the choice set, which are

SFD, SD, apartments, and other low-rise housing. It is assumed that all alternatives are available to all decision-makers, i.e., each developer is free to develop any type of housing.

The unknown parameters of the SMNL model were obtained directly by a maximum likelihood estimation of the log-likelihood function. Variables representing choice attributes and socioeconomic characteristics entered utility functions in generic or alternative-specific form. 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. These variables are used as proxies to present current housing stock and other location factors that affect the future housing developments in the same neighbourhood. 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.

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

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

As discussed earlier, in order to account for heterogeneity, a spatial mixed logit model (SML) is developed (see Equation 8). In random utility models, heterogeneity can be accounted for by allowing certain parameters of the utility function to differ across decision-makers. It has been shown that random parameter formulation can significantly improve both the explanatory power of models and the precision of parameter estimates.

SML model specification is similar to that of MNL and SMNL models, except that the parameters can vary in population rather than be the same for each decision-maker. As mentioned earlier, these parameters cannot be estimated analytically and must therefore be simulated for estimation purposes. In this study 1000 repetitions are used to estimate the unconditional probability by simulation. This will improve the accuracy of the simulation of individual log-likelihood functions and will reduce simulation variance of the maximum simulated log-likelihood estimator.

Two important aspects of modeling strategy that need to be considered before estimating a SML model are identifying parameters with and without heterogeneity as well as the assumption regarding the distribution of each of the random coefficients. These two must be selected based on prior information, theoretical considerations, or some other criteria. Random parameters in this study are estimated as normally distributed parameters in order to allow them to assume either negative or positive values. The observed attributes of the choices (explanatory variables) and their unobserved attributes (alternative specific con-

TABLE 2 Estimation Results and Comparison of Standard Multinomial logit (MNL) Spatial Multinomial Logit (SMNL) and Spatial Mixed Logit (SML) Models

Variable	Alt.[1]	MNL		SMNL		SML	
		Parameter	t-stat	Parameter	t-stat	Parameter	t-stat
Price	All	0.148	1.615	0.132	1.678	1.417	2.211
Std. Dev. of Price		--	--	--	--	1.664	4.906
Development Charge	D	-0.140	-2.715	-0.175	-2.725	--	--
	S	-0.167	-2.897	-0.173	-2.981	--	--
	O	-0.204	-3.257	-0.219	-3.480	--	--
	A	-0.288	-3.231	-0.301	-3.381	--	--
Intersection Density	D, S	-0.341	-5.359	-0.324	-5.947	-1.204	-6.849
School Accessibility	D	0.106	4.362	0.088	3.182	0.118	5.424
	S	0.106	4.362	0.091	3.239	0.169	2.691
Employment Accessibility	D	-0.084	-4.418	-0.075	-3.447	--	--
	S	-0.074	-3.858	-0.064	-2.900	-0.083	-1.610
	A	0.049	8.195	0.040	6.711	0.087	8.350
Inventory	A	-0.005	-3.296	-0.005	-3.614	-0.010	-3.994
Alternative Specific Constant	D	4.572	5.821	4.352	5.802	6.347	4.124
Std. Dev. of Detached Con.		--	--	--	--	2.986	1.848
Alternative Specific Constant	S	4.001	4.503	3.081	3.683	4.909	4.788
Std. Dev. of Semi-Det. Con.		--	--	--	--	1.773	1.540
Alternative Specific Constant	O	4.731	6.216	3.937	5.448	4.683	6.656
Std. Dev. of Others Con.		--	--	--	--	2.793	3.037
λ	D, S O, A	--	--	0.467	4.412	2.455	5.455
Std. Dev. of parameter λ		--	--	--	--	2.135	2.316
	D, S, O, A	--	--	2.663	2.922	2.663	fixed
Number of observations		1384.00		1384.00		1384.00	
Log-likelihood at zero		-1918.63		-1918.63		-1918.63	
Log-likeli. const.ant-only model		-1832.11		-1832.11		-1832.11	
Log-likelihood at convergence		-1478.77		-1453.48		-1291.44	
Log-likelihood ratio		0.226		0.242		0.324	

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

stants) were introduced as random parameters. Furthermore, the parameter of the spatial dependence term, λ, is assumed to be random, allowing heterogeneity across homebuilders with respect to the neighbourhood influence. Table 2 reports the results of SML model and its comparison with MNL and SMNL models to assess the importance of the parameter heterogeneity in these models.

Analysis of the Results

All parameters of the MNL model are statistically significant at the 95 % degree of confidence or better. Adding the spatial dependence term improves the overall goodness of fit of the model. The standard MNL model has an adjusted log-likelihood ratio of 0.226 when comparing the log-likelihood at zero and the log-likelihood at convergence. The constants alone contribute 0.045 of the 0.226, suggesting that the attributes in the utility expressions play an important role in explaining the housing type choice of new homebuilders.

The SMNL model with the spatial dependency term provides a better model fit over MNL model. The log-likelihood function value increased to -1453.48 and the log-likelihood ratio for the spatial logit model improved to 0.242. This confirms the robustness of the spatial logit model formulation and verifies the importance of the spatial dependency factors in explaining homebuilders' housing type choices.

The adjusted log-likelihood ratio of the SML model increased by 34 % to 0.324 in relation to the value for the SMNL model. This presents a significant improvement in the model fit confirming the improvement of the explanatory power of the model as a result of incorporating unobserved preference heterogeneity. The estimated standard deviations of the random parameters of price, λ, and alternative specific constants returned significant t-statistics, which indicated that they were statistically different from zero, and confirmed that parameters indeed were not consistent across the entire population. Results of the model strongly imply that heterogeneity is a significant factor in the model.

The signs of all utility parameters seem to be correct and unambiguous. The positive sign for the parameter on housing price provides evidence in favour of the hypothesis that homebuilders tend to be interested in building housing units that are likely to be sold for a higher price, which would increase their profit margins. Additionally, it can be hypothesized that the higher development taxes for a particular housing type will result in a lower choice probability of that particular housing type. As expected, the municipal development charge variable generated parameters with negative signs. This indicates that the development charge is negatively associated with the choice since it enters the model as a 'cost' variable. The development charge variable was not found to be a significant variable in SML model.

The model indicates that within the developed parts of the urban area, where street networks are highly developed and the intersection density is higher, homebuilders are likely to opt for building apartments, while the chance of building SFD and SD houses is lower. The parameters for SFD and SD 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.

Homebuilders respond to taste preferences and other needs of households by supplying types of housing that best meet the needs of households who are most likely to occupy those dwellings. For instance, households with children would prefer to locate in neighbourhoods with higher school accessibility. In addition,

households with children, because of their large size often occupy larger housing units. The statistically significant positive coefficients for the variable school accessibility in the utility functions of SFD and SD units capture the afore-mentioned dynamics.

The employment accessibility index has been used as a predictor in the model. The Central Business Direct (CBD) in Toronto has the single highest concentration of jobs in Toronto and is also very well-served by public transit. The employment accessibility index uses a gravity-type specification and therefore it assumes a higher accessibility values for areas in proximity of the CBD. The combination of high employment concentration and accessibility by rapid transit results in very high land values in and around the CBD. Therefore, the employment accessibility index also serves as a proxy for higher land values. The positive coefficient for employment accessibility in the utility function of apartments offers empirical evidence for the fact that in the areas with high land value, developers are more likely to develop apartments (condominiums), which use the minimum amount of land per housing unit. Similarly, the negative coefficients for the employment accessibility in the utility functions for SFD and SD housing suggests that homebuilders are less likely to develop housing developments that are land intensive.

The parameter of the inventory of residential units in the utility function for apartments is negative in the model. The large number of housing units undergoing the approval process in a zone suggest that there are large tracts of developable land in that zone, making it more attractive to develop SFD or SD housing and less attractive for apartment type developments.

The spatial dependency factor (λ) is introduced as a generic variable in the model. The positive sign of λ indicates that the existence of similar housing type in adjacent neighbourhoods increases the likelihood of the construction of the same housing type. This implies that homebuilders are influenced in their decision to build a particular type of housing by the decisions of other homebuilders in the vicinity.

Figure 1 presents the magnitude of spatial correlation approximated as a distance-decay curve, which has been specified as a function of the spatial parameters λ and γ. The curve indicates that development projects within a 4-km (2.5 miles) buffer will have a direct impact on the choice of the project type of homebuilders. Highly significant t-statistics for the parameter and its standard deviation suggest that neighbourhood effects are instrumental in determining housing type choices of homebuilders.

Conclusions

This paper presents the housing type choices of homebuilders while accounting for spatial dependencies and heterogeneity in taste parameters. The paper also presents derivation and development of a random parameter discrete choice model. Over the past few years, a relatively small body of research has attempted to capture the spatial and temporal dependencies across decision-makers and alternatives. While
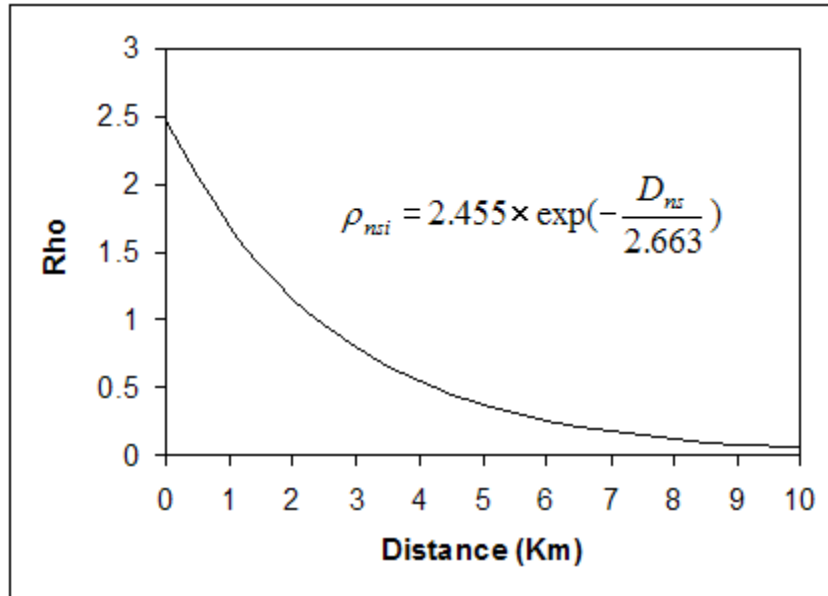
FIGURE 1 Distance-Decay Curve

temporal dependencies are often considered in dynamic models, there has been relatively little work in the literature on incorporating spatial dependencies into qualitative dependent variables and discrete choice models.

The basic idea presented in this paper is that homebuilders influence the housing-type choices of other builders who may choose to build new homes in close spatial proximity, resulting in a correlated choice behaviour over space. In this paper, spatial dependency terms are implemented in both standard and mixed multinomial logit frameworks. The results show that the spatial terms are statistically significant in the model and improve the model fit. Additionally, the model captures interactions between housing type, choice behaviour, and the existing land-use and accessibility. Further improvement to the model presented here can be offered by integrating the current SML model with a mixed GEV model. This will provide the opportunity to simultaneously account for spatial dependencies across decision-makers as well as unobserved spatial correlation across alternatives. Other improvement includes accounting for the effects of endogeneity in the spatial discrete choice model.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht, NL: Kluwer Academic.

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*. Norwell, USA: Kluwer Academic Publishers, 5-33.

Ben-Akiva, M. and D. Bolduc. 1996. "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure". Proceedings of the 3rd International Choice Symposium. New York: Columbia University, June 14-17.

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

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*. Berlin: Springer Verlag, 169-196.

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*. Oxford, UK: Elsevier Science, 381-414.

_____. 2000. "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling". *Transportation Science*, 34: 228-238.

_____. 1997. "An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel". *Transportation Science*, 31: 34-48.

Bhat, C.R. and J. Guo. 2003. "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling". Proceedings of the 82nd Annual Meeting of Transportation Research Board. Washington, D.C., January 13.

Boots, B.N. and P.S. Kanaroglou. 1998. "Incorporating the Effect of Spatial Structure in Discrete Choices Models of Migration". *Journal of Regional Science*, 28: 495-507.

Brownstone, D., D. Bunch and K. Train. 2000. "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuelled Vehicles". *Transportation Research B*, 34: 315-338.

Case, A. 1992. "Neighborhood Influence and Technological Change". *Regional Science and Urban Economics*, 22: 491-508.

Data Management Group. 1997. *TTS Version 3: Data Guide*. Toronto: Joint Program in Transportation, University of Toronto.

Dubin, R.A. 1992. "Spatial Autocorrelation and Neighborhood Quality". *Regional Science and Urban Economics*, 22: 433-452.

_____. 1995. "Estimating Logit Models with Spatial Dependence", in L. Anselin and R. Florax (eds.). *New Directions in Spatial Econometrics*. Heidelberg: Springer-Verlag, 229-242.

_____. 1998. "Spatial Autocorrelation: A Primer". *Journal of Housing Economics*, 7: 304-327.

Haider, M. 2003. *Spatio-temporal Modelling of Housing Starts in the Greater Toronto Area.* PhD thesis. Toronto: University of Toronto, Department of Civil Engineering.

Haider, M. and Miller, E.J. 2004. "Modeling Location Choices of Housing Builders in the Greater Toronto Area". *Transportation Research Record*, 1898: 148-156.

Louviere, J.J., D.A. Hensher, and J.D. Swait. 2000. *Stated Choice Methods, Analysis and Applications.* Cambridge, UK: Cambridge University Press.

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*. Heidelberg: Springer-Verlag, 189-228.

Mohammadian, A., and P. Kanaroglou. 2003. "Applications of Spatial Multinomial Logit Model to Transportation Planning". Proceedings of the 10th International Conference on Travel Behaviour Research. Lucerne, August 10-14.

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.

Revelt, D. and K. Train. 1998. "Incentives for Appliance Efficiency in a Competitive Energy Environment: Random Parameter Logit Models of Households' Choices". *Review of Economics and Statistics*, 80:.647-57.

Train, K. 1998. "Recreation Demand Models with Taste Differences Over People". *Land Economics*, 74: 230- 239.