Modeling Spatial Accessibility within Discrete Choice Framework

John M. Ulimwengu and Xiaoqi Guo

Authors are Graduate Research Associates in the Department of Agricultural, Environmental, and Development Economics, Ohio State University

Contact Information:

Department of Agricultural, Environmental, and Development Economics, Ohio State University; 2120 Fyffe Road, Columbus, OH, 43210

Ulimwengu.1@osu.edu and guo.63@osu.edu

Selected Paper prepared for the presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4, 2004

1. Introduction

Traditionally, accessibility has been conceptualized as the proximity of one location to other specified locations (Kwan and Weber, 2003). Based on cumulative opportunities and gravity concepts, spatial accessibility has been applied in various domains such as mixed-use core (MUC) designs (Limanond and Niemeier, 2003), neighborhood spatial accessibility to urban amenities (Hewko et al., 2002), housing developments (Geertman and Ritsema Van Eck, 1995), and accessibility of primary health care (Guagliardo, 2004). In the literature, definitions, measures and applications of spatial accessibility can be classified into three categories (Wee et al., 2001): infrastructures related, activities related and mixed measures. The first category focuses on the characteristics of infrastructure (for example, speeds on motorways). The second cluster is related to activities, such as living, working, recreating and shopping. It deals with the number of activities reachable within certain travel times or distances. The last category includes both infrastructure and activities characteristics.

In a more fundamental way, accessibility is concerned with the opportunity that an individual at a given location possesses to participate in a particular activity or set of activities. This definition captures the main feature of the concept of accessibility, which incorporates the underlying notions of spatial accessibility as well as social affordability. Indeed, accessibility comprises both physical and socio-demographic aspects. Accessibility measures based on traditional models, cumulative opportunities and gravity models, consider only variables defining distance between locations leaving out individual and spatial characteristics. Such measures implicitly assume that both individual and spatial characteristics are constant over individuals and geographical locations or irrelevant to the determination of accessibility. Obviously, such assumption has no theoretical or empirical justification. The presence of a hospital close to Mr.

Smith's residence does not necessarily guarantee that he will have access to health care from this particular hospital whenever he needs it; indeed, it may require health insurance which might be function of his income, and income itself may depend on his education, gender, race and probably the state of local or regional economic where Mr. Smith lives. Kwan and Weber (2003) agree that gravity-based and cumulative-opportunity measures are useful for identifying changes in the accessibility of different locations; they are also helpful for addressing issues of accessibility within transportation or information networks. However, these traditional measures are less suitable for understanding individual experiences due to recent changes in four areas: (a) the processes that shape urban form and contemporary urbanism; (b) the complexities of and individual difference in human spatial behavior; (c) the availability of new technologies and data for modeling individual accessibility; and (d) the increasing importance of information and communication technologies in people's lifestyle.

To address some of the weaknesses of traditional measures, models from random utility theory have been developed. The random utility approach corrects the lack of individual's involvement in gravity-based and cumulative-opportunity measures by explicitly introducing a decision process through utility maximization. The random utility approach as applied in the accessibility literature relies heavily on multinomial logit (MNL) (MacFadden, 1974) which, despite its closed form solution and readiness to interpret, does not account for possible spatial correlation. Indeed, the assumptions of independently and identically distributed (IID) random components imbedded in the multinomial logit model are no longer relevant when utilities from different spatial units are more likely to be correlated. Moreover, in MNL models the responsiveness to attributes of alternatives across individuals is assumed to be homogeneous after controlling for observed characteristics; a manifestation of the Independent from Irrelevant

Alternatives (IIA) property of the multinomial logit model. To improve accessibility measures from random utility theory, in this paper, we introduce frameworks that relax both independence and identity assumptions of the MNL models as well as unobserved response homogeneity assumption.

In the second section we discuss the relaxation of assumptions under which the MNL models are built; the third section presents the specifications of Multinomial Logit, Mixed Multinomial Logit and Nested Logit models; an application of these models for the derivation of spatial accessibility measures is presented in section four; and concluding remarks are exposed in the last section.

2. Relaxing MNL assumptions

As mentioned above, the MNL models are built under the assumptions of independently and identically distributed random components of the utility function. Such assumptions do not account for possible spatial dependence that is more likely to occur in discrete choice problems involving spatial units. The idea of spatial dependence finds its roots in the Tobler's (1979) "first law of geography" stating that "everything is related to everything else, but closer things more so," implying spatial dependence to be the rule rather than exception. As pointed out by Anselin (2002), inclusion of spatial dependence in applied models comes either from a formal specification of spatial interaction (see Brueckner, 2003) in an economic model or specificity of data exhibiting spatial dependence pattern. Failing to account for spatial dependence will result in biased estimates and incorrect predictions (Koppelman and Wen, 2000).

The most known relaxation of the independence assumption of the MNL model is the nested logit (NL) model, allowing for dependence between utilities of pairs of alternatives in the same groups (McFadden, 1978; Daly and Zachary, 1978). In the NL models the relative probability of two alternatives belonging to the same net is still independent of all the other alternatives. Furthermore, when the two alternatives are not in same nest, their probability ratio is independent of all alternatives in all the other nests except the two they belong to, resulting in the so called Independence of Irrelative Group (IIG) property. Other relaxations of the independence assumption of the MNL model are found in ordered generalized extreme value (OGEV) model (Small, 1987), the paired combinatorial logit (PCL) model (Chu, 1989; Koppelman and Wen, 2000), cross-nested logit (CNL) model (Vovsha, 1997), the multinomial logit-ordered GEV (MNL-OGEV) model (Bhat, 1998), and the product differentiation logit (PDL) model model (Bresnahan et al., 1997), all derived from McFadden's generalized extreme value (GEV). Mixed multinomial logit (MMNL) class of models has been developed to account for the unobserved response homogeneity (Revelt and Train, 1998). To relax both IID assumption and unobserved homogeneity while avoiding computational difficulties associated with MMNL models, Bhat and Guo (2004) propose to superimpose a mixing distribution over the GEV structure; the resulting model is called the mixed spatially correlated logit (MSCL) model. In this paper we compare accessibility measures derived from the MNL, NL, and MMNL models as described below.

3. Derivation of the Functional Forms of MNL, NL and MMNL

To derive functional forms for the MNL, NL and MMNL models, we assume that individual n has to choose over a set of I spatial units (i=1, 2,…,I). Omitting the subscript n for the decision maker and setting $Y_j \equiv \alpha_{i,ij} \exp(V_j)$[1], we consider a function, $G = G(Y_1,..., Y_J)$ with $G_i = \partial G/\partial Y_i$. It is easy to see that:

1. $G \geq 0$ for all positive[2] values of $Y_j \forall j$.

2. G is homogeneous of degree one in Y_j.

3. $G \rightarrow \infty$ as $Y_j \rightarrow \infty$ for any j.

4. The cross partial derivatives of G change signs in a particular way. That is, $G_i \geq 0$ for all i, $G_{ij} = \partial G_i / \partial Y_j \leq 0$ for all $j \neq i$, $G_{ijk} = \partial G_{ij} / \partial Y_k \geq 0$ for any distinct i, j and so on for higher-order cross-partials. Therefore, discrete choice models can be derived based upon G. Thus,

$$P_i = \frac{Y_i G_i}{G} \tag{3}$$

is the probability for a discrete choice model that is consistent with utility maximization. In addition, we assume that the random component (ε_i) of utility function follows an extreme-value distribution

3.1. Multinomial Logit

The MNL model is obtained, with $G = \sum_{j=1}^{J} Y_j$, as follows:

$$P_i = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)} \tag{4}$$

---

[1] V is the observed component of the utility function and depends on exogenous variables (X): $V = X\beta$

[2] By construction Y_j is necessarily positive.

3.2. Mixed Multinomial Logit

As pointed out by Train (2002), mixed logit probabilities are integrals of standard logit probabilities evaluated at parameters (β). Explicitly, the probability in mixed logit is given by

$$P_i = \int \frac{\exp(V_i(\beta))}{\sum_{j=1}^{J} \exp(V_j(\beta))} f(\beta)d\beta \qquad (5)$$

where $V_j(\beta)$ is the observed component of the utility function from alternative j, and $f(\beta)$ is a density function. If instead the homogeneity response assumption (constant β) is correct, the MMNL collapses to MNL.

3.3.Nested Logit

The nested logit family is derived by choosing $G = \sum_{l=1}^{K}\left(\sum_{j\in B_l} Y_j^{1/\lambda_l}\right)^{\lambda_l}$ where J alternatives are partitioned into K nests labeled $\mathrm{B}_1,...,\mathrm{B}_\mathrm{K}$ and $0 < \lambda_l \leq 1$ for each l. These nets are such that for any two alternatives in the same nest, the IIA property holds; but in general does not hold for alternatives in different nests. The functional form for the mixed logit probabilities is given by

$$P_i = \frac{Y_i G_i}{G} = \frac{\exp(V_i/\lambda_k)\left(\sum_{j\in B_k}\exp(V_j/\lambda_l)\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l}\exp(V_j/\lambda_l)\right)^{\lambda_l}} \qquad (6)$$

If $\lambda_k = 1\,\forall k$, the NL is equivalent to the MNL.

Following McFadden (1981) expected maximum utilities for MNL, MMNL and NL are computed as follows:

$$E_n^{MNL} = \ln\left[\sum_{j=1}^{J}\exp(V_j)\right] \qquad (7)$$

$$E_n^{MMNL} = \int \ln\left[\sum_{j=1}^{J} \exp(V_j(\beta))\right] f(\beta) d\beta \qquad (8)$$

$$E_n^{NL} = \ln\left[\sum_{l=1}^{K} \left(\sum\nolimits_{j \in B_l} \exp(V_j / \lambda_l)\right)^{\lambda_l}\right] \qquad (9)$$

4. Measuring Accessibility

As pointed out by Handy and Niemeier (1997), the cumulative opportunities measures are the simplest measures of accessibility; they count the number of opportunities reached within a given travel time (or distance). They provide some idea of the set of choices available to residents, for example, in terms of housing units they can choose from. The second type of accessibility comprises the gravity-based measures that weight opportunities, usually the quantity of an activity as measured by employment, impendence, generally a function of travel time or travel cost (Handy and Niemeier, 1997). Under this approach, accessibility, A_i, for residents of location i is measured as

$$A_i = \sum_j a_j f(t_{ij}), \qquad (10)$$

where a_j is the activity in location j, t_{ij} is travel time, distance, or cost from location i to location j, and $f(t_{ij})$ is an impedance function which can take different functional forms.

The last class of accessibility measures is derives from the random utility approach. In this case, it is assumed that a resident assigns utility to each destination/location choice in some specified choice set and then selects the alternative that maximizes his utility. In this paper, accessibility measures are defined by the denominators of equations (4), (5) and (6) respectively for the MNL, MMNL and the NL models. These accessibility measures are equivalent to the maximum expected utility derived in section 3. Explicitly, accessibility, A_n, for an individual, *n*, is measured by:

$$A_n^{MNL} = \ln\left[\sum_{j=1}^{J} \exp(\hat{V}_j)\right] \tag{11}$$

$$A_n^{MMNL} = \int \ln\left[\sum_{j=1}^{J} \exp(\hat{V}_j(\hat{\beta}))\right] f(\beta) d\beta \tag{12}$$

$$A_n^{NL} = \ln\left[\sum_{l=1}^{K} \left(\sum\nolimits_{j\in B_l} \exp(\hat{V}_j / \hat{\lambda}_l)\right)^{\hat{\lambda}_l}\right] \tag{13}$$

5. Application

A discrete-choice model of residential location is estimated using data from the record of residential location and survey on 824 homeowners in Franklin County, Columbus, Ohio in 1995. The dependent variable is the zero-one indicator of residential location the homeowner chooses from 17 available school districts. The independent variables include household income; leisure-security expenditures ratio as a proxy for the quality of public goods in the district; population density; percentage of residents with college degree; commuting time from the central node of each district to downtown Columbus as proxy of physical accessibility to major employment and entertainment destinations; number of retailed business per capita in each district; and housing units built before 1970 to proxy the quality of housing stock. The results of estimated parameters are shown in Table 1 and the average estimated accessibility values for three different income groups are presented in Table 2.

Table 1: Estimation results for the MNL, NL and MMNL[3]

	MULTINOMIAL LOGIT		NESTED LOGIT		MIXED LOGIT	
Variables	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
% of house built before 1970						
Mean	-9.808	-9.50	-7.845	-7.16	-11.896	--9.99
Standard deviation*	0.000	-	0.000	-	0.197	0.08
Household size						
Mean	3.381	6.59	3.235	6.169	4.009	7.13
Standard deviation*	0.000	-	0.000	-	0.148	0.17
Population density						
Mean	-81.795	-3.01	-105.044	-3.95	-134.157	-2.54
Standard deviation*	0.000	-	0.000	-	3.464	0.12
% of college graduate						
Mean	-0.826	-2.35	-0.093	-.0.17	-1.583	-2.65
Standard deviation*	0.000	-	0.000	-	1.827	0.83
% commuting time to Downtown						
Mean	-0.164	-10.82	-0.143	-7.89	-0.194	-9.29
Standard deviation*	0.000	-	0.000	-	0.128	3.19
Leisure/Security expenditure ratio						
Mean	-0.083	-2.291	-0.078	-2.09	-0.901	-1.25
Standard deviation*	0.000	-	0.000	-	0.809	1.31
$\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_2 = \hat{\lambda}_4$**	1.000	-	0.890	7.23	-	-
$\hat{\lambda}_5$**	1.000	-	0.627	3.50	-	-
# Observations	824		824		824	
Log likelihood at convergence	-2054.772		-2048.465		-2067.099	
Estimation method	Maximum Likelihood		Full Information Maximum Likelihood		Maximum Likelihood	

* The standard deviations are implicitly constrained to 0 in the MNL and NL model.
** The index of similarity between alternatives is implicitly constrained to 1 in MNL.

As shown in table 2, a misspecification of the distribution of the error terms and that of the coefficients (fixed or random) may lead to contradictory results. Indeed, our results suggest that the MMNL systematically underestimates accessibility measures where either MNL or NL is the true models. Similarly, the MNL underestimate spatial accessibility whenever the NL is believed to be the appropriate; specially, when indexes of similarity between alternatives are

[3] 17 school districts specific-income coefficients have been estimated along with estimates reported in table 1. These marginal utilities of income were used to compute individual accessibility measures. Table 3 combined with Map 1 present different nest built from geographical location of school districts.

significantly different from one as is the case in our illustrative study. However, the most intriguing result comes from the ranking of accessibility measures across income groups; whereas measures from NL grant low-income households with low level of accessibility, high-income households receive low level accessibility under MNL and MMNL frameworks. This result implies that a misspecification is more likely to induce inefficient policy measures. For example, assuming a marginal utility of income of 1 everywhere, under the "farm land preservation" scheme, the government will have to compensate more low-income households if farm land accessibility is derived from MNL and the reverse will happen under the NL model. The lesson here is that the simplicity of a model structure often presented as the choice criteria must always be weighed against the risk of "corrupting" the decision process.

Table 2: Spatial Accessibility Measures

Income groups	NL	MNL	MMNL
≤ $100,000	6.492937	5.830051	3.915493
$100,001-$1,800,000	6.662112	5.750936	4.162848
> $1,800,000	7.079317	5.730813	4.205952

Obviously, accessibility measures reported in table 2 do not have an economically sounds interpretations. Small and Rosen (1981) demonstrated that the conventional methods of applied welfares economics can be used in the case of stochastic utility models such as discrete choice models. Following their results, in order to obtain an economically sound interpretation of the derived accessibility measure, accessibility measures can be used to compute change in consumer surplus. Then, using marginal utility of income, the consumer surplus can be transformed into compensating variation expressed in monetary terms; thus, readily interpretable and usable for comparisons purpose. Handy and Niemeier (1997) interpret the resulting measure as the accessibility worth or the amount someone must be compensated by after an endogenous shock (change in policy) that reduces accessibility in order to be as better-off as before the shock.

6. Concluding remarks

The main objective of our paper was to highlight the shortcomings of traditional accessibility measures and provide some appropriate methodological suggestions for their improvement. Traditional measures derived from cumulative opportunities and gravity models focus on physical proximity leaving out individual and spatial attributes as potential explanatory variable. It is clear that an individual may be physically close to, say, a hospital yet unable to purchase its services because of lack of health insurance or income adequate to cover the costs. The correction brought by random utility theory relies mainly on MNL models under IID and individual response homogeneity assumptions that often do not hold in case of choices involving spatial units. In this paper we briefly present the process of relaxing MNL assumptions. Using the MNL, NL, and MMNL models we derive related accessibility measures. The application of MNL, NL and MMNL on choice model of residential location underlines possible consequences of a misspecification of the distribution of the error term and that of model parameters. The results clearly suggest that a decision process can be corrupted, and therefore lead to erroneous policy measures because of model misspecification. Moreover, the simplicity of a model structure, though appealing, does not necessarily guarantee the best outcome in terms policy design.

References

Anselin, L. (2002). “Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models,” *Agricultural Economics* 17:3, 247-267.

Bayoh, Isaac, Irwin, E.G., and Haab, T. (2002). "Estimating the Determinants of Household Residential Location Choice Using a Multinomial, Discrete Choice Model." Manuscript.

Bhat, C. (1998). “Accommodating variations in responsiveness to level-of-service variables in travel mode choice models,” *Transportation Research A* 32, 455–507.

Bhat, C. and J. Guo (2004). “A Mixed spatially correlated logit model: formulation and application to residential choice modeling,” *Transportation Research B* 38, 147-168.

Bresnahan, T.F., S. Stern, and M. Trajtenberg (1997). “Market segmentation and the sources of rents from innovation: personal computers in the late 1980s,” *RAND Journal of Economics* 28, 17-44.

Brueckner, J. K. (2003). “Strategic Interactions among Governments: An Overview of Empirical Studies,” *International Regional Science Review* 26:2, 175-188.

Chu, C. (1989). “A paired combinational logit model for travel demand analysis,” *Proceedings of Fifth World Conference on Transportation Research* 4, 295–309.

Daly, A. and S. Zachary (1978). “Improved multiple choice models,” in D. Hensher and in M. Dalvi, eds., *Determinants of Travel Choice*, Saxon House, Sussex.

Geertman, S.C.M. and J.R. Ritsema van Eck. (1995). “GIS and models of accessibility potential: an application in planning,” *International Journal of Geographical Information Systems* 9:1,67-80.

Guagliardo, M.F. (2004). “Spatial accessibility of primary care: concepts, methods and challenges,” *International Journal of Health Geographics* 3:3

Handy, S. and D.A. Niemeier (1997). “Measuring Accessibility: An Exploration of Issues and Alternatives,” Environment and Planning A 29, 1175-119.

Hewko, J., K.E. Smoyer-Tomic, and M.J.Hodgson (2002). “Measuring Neighbourhood Spatial Accessibility to Urban Amenities: Does Aggregation Error Matter?” *Environment and Planning A* 34:7, 1185-1206.

Kawabata, M. (2003). “ A GIS-based Analysis of Jobs, Workers and Job Access in Tokyo,” Center for Spatial Information Science, CSIS *Discussion Paper* No. 57, University of Tokyo.

Koppelman, F. and C. Wen (2000). “The paired combination logit model: Properties, estimation and application,” *Transportation Research B* 34, 75–89.

Kwan, M.P. and J. Weber. (2003). “Individual Accessibility Revisited: Implications for Geographical Analysis in the Twenty-first Century,” Geographical Analysis, 35:4.

Limanond, T. and D.A. Niemeier. (2003). “Accessibility and mode-destination choice decisions: exploring travel in three neighborhoods in Puget Sound, WA,” *Environment and Planning B: Planning and Design* 30, 219-238.

McFadden, D. (1974). “Conditional logit analysis of qualitative choice behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105–142.

McFadden, D. (1978). “Modeling the choice of residential location,” in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75–96.

McFadden, D. (1981), “Econometric models of probabilistic choice,” in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge, Massachusette; M.I.T. Press, 198-272.

Revelt, D. and K.Train (1998). “Mixed logit with repeated choices,” *Review of Economics and Statistics* 80, 647–657.

Small, K. (1987). “A discrete choice model for ordered alternatives,” *Econometrica* 55, 409–424.

Small, K. and H. Rosen (1981). “Applied welfare economics of discrete choice models,” *Econometrica* 49, 105–130.

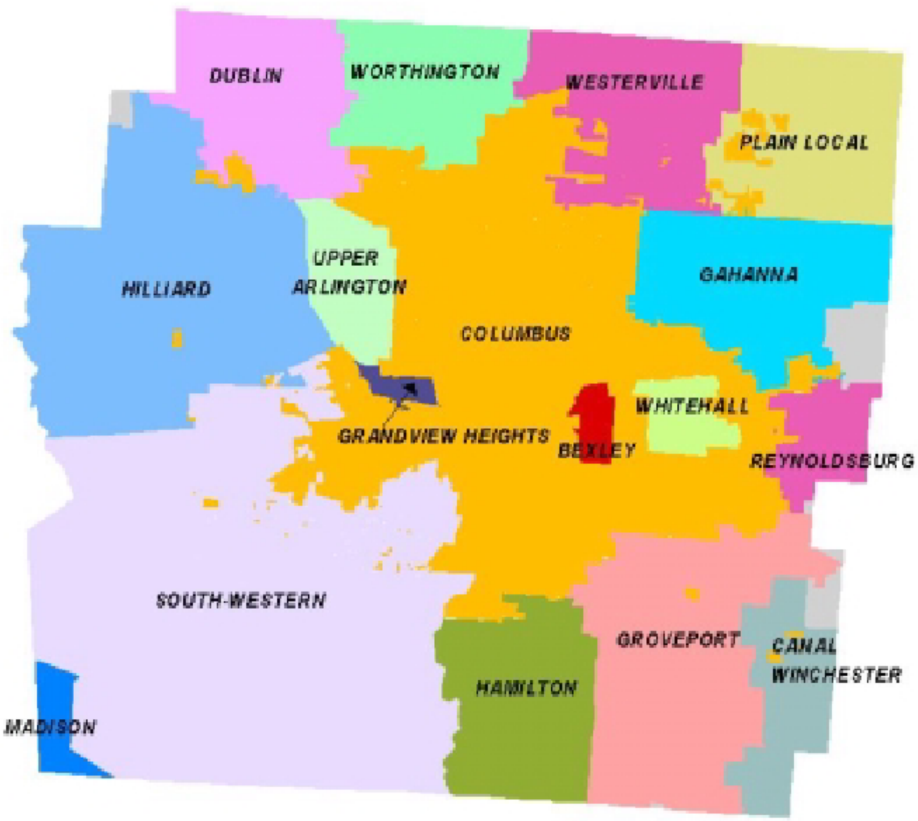
Tobler, W. (1979). “Cellular Geography,” in S. Gale and G. Olsson (eds.), *Philosophy in Geography*, Reidel, Dordrecht, 379-386.

Train, K. (2002). Discrete Choice Methods with Simulation. Cambridge University Press, New York.

Vovsha, P. (1997). “The cross-nested logit model: Application to mode choice in the Tel Aviv metropolitan area,” Conference Presentation, 76th Transportation Research Board Meetings, Washington, DC.

Wee, B., M. Hagoort and J.A. Annema. (2001). “Accessibility measures with competition,” *Journal of Transport Geography* 9, 199-208.

Map 1: School District Choice Set
Franklin County, Ohio



Source: Bayoh et al. (2002)

Table 2: School district nets

Nets	School districts
1. Northeast	Westerville
	Plain local
	Gahana
2. Northwest	Worthington
	Dublin
	Hilliard
	Upper Arlington
3. Southeast	Canal Win
	Grove Mad
	Hamilton
	Reynolds
	Whitehall
4. Southwest	South West
	Madison
5. Central	Bexley
	Columbus
	Grandview