

THE IMPACT OF SPACE ON THE APPLICATION OF DISCRETE CHOICE MODELS

Kingsley E. Haynes and A. Stewart Fotheringham*

I. Introduction

In spatial analysis, the 1980's could be characterized as the decade of discrete choice modeling and, more generally, categorized data analysis (Williams, 1977; Wilson, 1981; Wrigley, 1982; 1985; Wrigley and Longley, 1984). Subsequent to the pioneering work undertaken in economics, marketing and transportation, it was quickly recognized that most spatial decisions are discrete (for example, the selection of a city in which to live or a store at which to shop) and the discrete choice framework was adopted enthusiastically (Fotheringham and O'Kelly, 1989). One of the most popular statistical models for the analysis of discrete choices has been the multinomial logit model (MNL). There are good reasons for its popularity: the model is consistent with utility-maximizing behavior; the selection probabilities can be expressed in closed form; parameter estimation is straight forward and the general approach has been shown to be informationally consistent across a range of applications and forms (Haynes and Phillips, 1982).

Space, however, provides a much more complex background against which to model discrete choice than do the aspatial contexts in which the discrete choice modeling (DCM) framework was developed. In this paper we describe the added complexities space introduces into DCM and then discuss how they can be incorporated into the framework to produce more realistic spatial choice models. Our discussion is centered around the MNL model, because of its popularity, and on the role of the Independence from Irrelevant Alternatives (IIA) assumption which is shown to be a key factor in highlighting the differences between aspatial and spatial choice. We also demonstrate that in certain models where IIA is relaxed the subtleties of space are not captured sufficiently and that these subtleties can only be captured by developing models

*Dean of the Graduate School, Director of the Public Policy Institute and Professor of Public Affairs, Geography and Decision Sciences at George Mason University; and Professor of Geography and Research Scientist, National Center for Geographic Information and Analysis, State University of New York at Buffalo, respectively. The support of NCGIA under NSF grant #SES-8810917 is acknowledged.

from spatial theory. Some promising beginnings in this direction are reported.

The structure of the paper is as follows: The first section outlines the implications of IIA for modeling discrete choice processes with particular reference to the MNL. The next section focuses on the reliability of these assumptions in a spatial context. Attention then turns to two alternative paths for relaxing the IIA assumption through explicitly modeling alternative substitutability within the systematic component of utility or implicitly modeling alternative substitutability through the error structure. Throughout these latter two sections, we focus on the potential for introducing spatial effects into existing model frameworks.

II. Rationality and Independence

As an axiom of his models for rational probabilistic choice over a set of discrete alternatives, Luce (1959) assumed that the decision process could be characterized by independence from irrelevant alternatives (IIA). The IIA assumption can be stated as:

$$\frac{P(j|C')}{P(k|C')} = \frac{P(j|C)}{P(k|C)} \quad j, k \in C', \forall C' \subseteq C \quad (1)$$

where,

- C is the set of all alternatives available to the decision-maker;
- C' is a subset of C;
- P(j|C) is the probability of alternative j being chosen given that the choice set is C.

Equation (1) implies that the odds of any element j being chosen over k are independent of the size and composition of the choice set, i.e., the choice axiom constrains the odds to be constant over all possible choice sets, containing j and k.

A second component of Luce's approach is the assumption of rationality on the part of the decision-maker, i.e., the (discrete) alternative which yields the highest level of utility will be chosen. The utility assigned to an alternative is typically represented as (McFadden, 1974):

$$U_{ij} = V_{ij}(x_{ij}, s_i) + e_{ij}(x_{ij}, s_i) \quad (2)$$

where,

- x_{ij} is a vector of attributes of alternative j for individual i ;
 s_i is a vector of characteristics of individual i ;
 V_{ij} is a function measuring the effect on utility of the attributes x for the "representative" or "average" consumer with characteristics s ;
 e_{ij} is an individual-specific portion of utility.

When the individual portion of utility is stochastic, the alternative with the highest level of utility can only be predicted up to a probability:

$$P_i(j|C') = \Pr(U_{ij} - U_{ik} > 0) \quad \forall k \in C', j \neq k \quad (3)$$

$$= \Pr(e_{ij} - e_{ik} > V_{ik} - V_{ij}) \quad \forall k \in C', j \neq k$$

Tversky (1972) has shown that whenever the random components of utility are uncorrelated across alternatives, then the decision rule is simply scalable. This implies that the utilities associated with each alternative can be measured or represented on a single continuous scale, so that the choice problem can be solved by pairwise comparisons between the scaled utility measures, i.e., $(U_{ij} - U_{ik})$ in Equation (3). Tversky also shows that simple scalability implies order independence, i.e., an individual's preference ranking of any two alternatives is independent of the composition of the choice set. IIA is a special case of simple scalability - the odds constraint obviously ensures order independence. This means that the assumption of zero correlation between the error components across alternatives entails the IIA hypothesis of Equation (1). Thus, prediction of the odds of j being chosen over k are conditioned by the characteristics of those two alternatives and are independent of the characteristics of other (irrelevant) alternatives.

Specification of the selection probabilities for a random utility model proceeds by assuming a joint distribution for the e_{ij} 's. A convenient assumption is that the random utility components are independently and identically distributed according to the Type I Extreme Value distribution¹. This leads to the multinomial logit (MNL) model for selection probabilities:

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_k \exp(V_{ik})} \quad (4)$$

The MNL is particularly appealing because the selection probabilities, under the distributional assumption, are consistent with utility-maximization by the decision-maker. The MNL clearly satisfies the IIA restriction since the odds for any pair of alternatives are unaffected by the character-

istics of other alternatives, as can be seen by taking the odds ratio:

$$\frac{P_{ij}}{P_{ik}} = \frac{\exp(V_{ij})}{\exp(V_{ik})} \quad j \neq k, j=1, \dots, j, k=1, \dots, K \quad (5)$$

Typically, a utility maximization model entails some set of constraints on the choice variables. For example, in the classical economic model, substitutability between commodities arises because the individual maximizes utility subject to an income constraint. Without the income constraint, there would be no need to trade off one commodity for another. Thus, the characteristics of other commodities (their prices) enter into the decision rule (demand equation) which determines the optimal level of consumption of some commodity. Further, each commodity can be consumed at a non-zero level, subject to the nonnegativity constraints. By contrast, in the discrete choice context, the only constraint which is imposed is that only one alternative can be chosen - all other alternatives must be 'consumed' at a zero level. In that sense, all of the alternatives are equally substitutable in satisfying the constraint that one be chosen. However, a discrete choice model which assumes IIA can yield erroneous predictions if the elements of the choice set exhibit unequal substitutability.

Two possible strategies are available for remedying this situation: explicitly incorporating attributes of other alternatives into the observable component of the utility function or relaxing the assumption of simple scalability and zero error covariances. The applicability of these alternative approaches in spatial choice analysis will depend to a large extent on the particular structure of the choices in any given context.

III. Spatial Choice

Luce-type models, which assume IIA, have frequently been used in the analysis of spatial choice decision, both for analyses of aggregate data, or flows of people, and for micro-level analyses, which focus on individual choices. Applications of probability models of the form of Equation (5) to flow data are particularly prevalent in the area of destination choice, as, for example, the production-constrained gravity model which originated with Huff (1962) and which is discussed in a random utility setting by Shepherd (1978). Hertzvitz's (1983) migration study is an application to aggregate flow data, which explicitly invokes random utility theory and discusses some of the additional assumptions involved in using the micro-level choice theory for aggregate flows. The MNL has been used in a variety of contexts for micro-level studies, including migration (Linneman and Graves, 1983;

Mueller, 1982; Odland and Ellis, 1987); industrial location (Carlton, 1983; Hansen, 1987); choice of a recreational facility (Brown, 1979); choice of housing type (Quigley, 1976); residential location and housing type (Anas, 1984; Clark and Onaka, 1985; Onaka and Clark, 1983; Quigley, 1985); and choice of shopping destination and travel-demand related decisions (Domencich and McFadden, 1975).

Spatial choice sets, however, exhibit a number of characteristics not found typically in aspatial choice problems which can render the IIA assumption questionable. These characteristics can be summarized as follows:

Size

Spatial choice sets generally have a much larger number of alternatives than is found in aspatial contexts. Large choice sets are not a problem *per se*, as long as the IIA assumption holds. In fact, the IIA assumption would be particularly useful when there are a large number of alternatives due to its simplification of the choice process and computational tractability in the face of large choice sets. Further, when the choice set is very large, spatial choice models can be calibrated on a sample of alternatives, a procedure which can be justified by the IIA property (see Ben-Akiva and Lerman, 1985). However, it is also the case that, with large choice sets, the equal substitutability constraint may be more difficult to satisfy. Consider, for example, the problem of predicting destination choices of migrating households and suppose that the choice set is the 48 co-terminus states of the U. S. In that context, it is quite possible that, for a destination such as, say, North Dakota, a state such as South Dakota will be more equally substitutable than states such as, say, California or New York. An alternative way of viewing this is to consider that individuals have a limited capacity for storing and processing information. As a result, choice sets with large numbers of alternatives tend to be subdivided into clusters each of which has a manageable number of alternatives, that can be evaluated. This mental partitioning of large choice sets produces a situation where alternatives within the same subset have a greater degree of substitutability than do alternatives in different subsets.

Aggregation

The elements of spatial choice sets can often be characterized as groupings or aggregations of elemental alternatives, i.e., the precise points in space at which location decision processes are terminated. For example, the SMSA's, in Carlton's (1983) study were themselves the choice sets - a firm choosing a set of SMSA's. Thus, grouping of alternatives implies an explicit choice model for the aggregates and an implicit model for the elemental

alternatives within each grouping. The latter is a conditional model, since choice of an aggregate is assumed to precede choice of an elemental alternative.

The theory of aggregation of elemental alternatives (see, *inter alia*, McFadden (1984) and Ben-Akiva and Lerman (1985) suggests that grouping of alternatives can undermine the equal substitutability constraint in two ways. First, if alternatives are grouped, the expression for the systematic portion of utility of grouped alternatives should control for variations across grouped alternatives in the number of elemental alternatives within each grouping (i.e., the 'size' of each grouping) and also for the heterogeneity of the elemental alternatives). Exclusion of size and heterogeneity measures will, apart from introducing misspecification bias due to erroneously excluded variables, undermine IIA - for example, excluding a size measure when the groupings vary in this respect means that errors across grouped alternatives must be correlated. Typically, the degree of heterogeneity of elemental alternatives within groups and/or the number of elemental alternatives in each group is unknown and proxy variables have to be used. The use of proxy variables, however, introduced error into the random component of utility of grouped alternatives and, again, the assumption of IIA less tenable since it cannot be assumed, *a priori*, that such errors will be independent across grouped alternatives.

Second, equal substitutability will fail to hold if the random components of utility for the elemental alternatives are not independent, i.e., if the equal substitutability constraint does not hold for the elemental alternatives which comprise a particular grouped alternative. Thus, application of MNL to grouped alternatives entails the assumption that the choice between elemental alternatives can also be characterized by IIA. This holds even if size and heterogeneity measures are incorporated and/or unnecessary.

Dimensionality

Many spatial choice problems can be characterized as multi-dimensional. For example, the migration decision process can be broken down into the move/stay decision and the destination choice decision; intra-urban residential mobility involves both the choice of a neighborhood and the choice of a particular dwelling unit. Treating a multi-dimensional choice problem as a unidimensional choice problem can render the equal substitutability constraint untenable. To see this, consider the residential location problem and suppose that the utility function for individual *i* is specified as:

$$U_{ijk} = V_{ijk} + e_{ijk} \quad (6)$$

where,

j subscripts the neighborhoods of choice, $j=1, \dots, J$;
 k subscripts the housing units, k , in neighborhood j ,
 $k=1, \dots, M_j$.

The housing units in each neighborhood will share the unique characteristics of that neighborhood, such as access to the C.B.D., public service provision levels, and so on. Thus, (6) can be decomposed along the dimensions of choice as (see Ben-Akiva and Lerman, 1985):

$$U_{ijk} = (V_{ij} + e_{ij}) + (V_{ik} + e_{ik}) + e_{ijk}^* \quad (7)$$

Where,

V_{ij} and e_{ij} are the observable and unobservable neighborhood-level component of utility.

V_{ik} and e_{ik} are the observable and unobservable components of utility which are specific to the dwelling units in j .

and, e_{ijk}^* is a random component of utility shared by choice dimensions.

If the utility function can be decomposed as in (7) above, then, even if the separate elements of the random component are independent of each other within each alternative and are each independent across alternatives, the only way that equal substitutability can hold for the full choice set is if the e_{ij} 's have zero variance. This follows since the covariance between dwelling-specific random components which share a common neighborhood is equal to the variance of the neighborhood level random component which is common to both dwelling units:

$$E(e_{ijk} e_{ijl}) = e_{ij}^2 = \sigma_j^2 \quad k, l \in j, k \neq l \quad (8)$$

Multi-dimensional choice sets can arise in three ways. First, there are "naturally" dimensioned choice sets, as in the move/stay and destination choice dimensions in migration. Multiple dimensions can also arise when the observed attributes vary at different spatial scales. For example, in the housing choice problem considered above, the neighborhood-level attributes are constant within neighborhoods as can be seen from Equation 7. If there are components of utility which vary at different spatial scales, then the random component can, in principle, be decomposed as in (7) which leads to the potential IIA violation (8). Finally, multiple dimensions can arise when the choice set contains elements which share some qualitative aspect, as for example C.B.D. locations versus suburban locations. This sharing of qualitative aspects can, in principle, be facilitated within a MNL framework by means of dummy variables. However, such a strategy means that some component of utility will vary across sub-sets of the choice set while being constant within sub-sets, i.e., the utility function would resemble Equation 7.

Spatial Continuity

In any choice situation with a large number of alternatives, it seems clear that individuals partition the choice set into clusters of alternatives. It is usually easier to identify clusters of alternatives in aspatial choice, such as brands of decaffeinated coffee, sports cars, types of public transit, etc., than it is in spatial choice. This difference arises because, whereas in aspatial choice, the discriminating factor between clusters is discrete (an automobile is either a sports car or a non-sports car), the discriminating factor between clusters of spatial alternatives is usually space itself, which is continuous. The imposition of discrete boundaries in such an instance is subjective and rather arbitrary. Boundaries of spatial clusters are more likely to be 'fuzzy' (Zadeh, 1965; Gale, 1972; Pipkin, 1978) than discrete which causes problems in the application of choice models such as the nested logit that need an *a priori* definition of clusters. The successful application of such models in an aspatial context thus offers no guarantee of similar success in a spatial context.

The continuous nature of space affects choice in another way. In clusters of aspatial alternatives there is an implicit assumption of randomness in the arrangement of alternatives which within a cluster, are assumed to be equal substitutes for one another. In a spatial choice cluster, however, there is an ordering to the alternatives, resulting from the final locations of alternatives and the different spatial relationships that exist for each alternative. Consequently, the more two alternatives are separated by space, the less likely they are to be substitutes for each other. This again reflects the discrete nature of most aspatial relationships. Formally, if $S_{jj'}$ is defined as the degree of substitution between two alternatives, j' and j , the following relationships are hypothesized:

In spatial choice,

$$S_{jj'} = \begin{cases} 1 & \text{if } j' \text{ and } j \text{ are in the same cluster} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

In aspatial choice,

$$S_{jj'} = f(d_{jj'}) \quad (10)$$

where $f(\cdot)$ represents a continuous function and $d_{jj'}$ represents the spatial separation between j' and j .

It can be seen from Equations (9) and (10) that while there is transitivity between the substitution of aspatial choices, there is no guarantee of a similar transitivity in spatial choice. Consider, for example, a cluster of automobiles defined as Luxury cars consisting of a Rolls-Royce, a Jaguar and a Mercedes. Suppose that the Rolls-Royce and Jaguar are substitutes for one another and that the Jaguar and Mercedes are also substitutable. It then

follows that the Rolls-Royce and Mercedes are substitutes, or, at least, there is a high probability that they are. This same transitivity is much less likely to occur in a spatial setting. Consider three destinations, A, B and C, located in a cluster and that B lies midway between A and C. If A and B and B and C are substitutes, because of their greater spatial separation, there is no guarantee that A and C are substitutes.

Variation and Location

Consider again the situation where an individual is faced with making a choice from a large number of alternatives, so that the alternatives are partitioned into clusters and only the alternatives in one cluster are examined in detail. In aspatial choice sets, the set of alternatives forming each cluster can usually be assumed to be constant (for example, the perception of the difference between regular and decaffeinated coffee or between private and public transit, is unlikely to vary over space). This is a useful assumption to make in nested choice models (such as the nested logit) since only one hierarchy of choices needs to be identified for all individuals. Unfortunately, it is an assumption that is very unlikely to hold for spatial choice problems where the evidence from the literature on mental maps (Gould and White, 1974) strongly suggests that individuals in different locations have different perceptions of space. A migrant from New York, for example, is likely to partition space differently than a migrant from Arizona. The former's mental construct of opportunity sets is likely to contain a finer partitioning of opportunities in the North East, while the latter is likely to contain a finer partitioning of opportunities in the South West. The implication of this for spatial choice modelling, is that if a model is employed in which the partitioning of choice sets needs to be defined *a priori*, as in the nested logit model, in theory a different partition would have to be employed for each individual, or at least, every origin.

The preceding discussion has highlighted some of the problems which arise in spatial choice modeling and which can render the IIA assumption of equal substitutability untenable. The next two sections discuss alternative approaches, which have been formulated for modeling unequal substitutability.

IV. Systematic Component Substitutability

It is possible to sidestep, at least to some degree, the IIA problem by altering the set of regressors that are used in the functional form for V_{ij} . For example, McFadden's (1981) universal logit model allows attributes of all alternatives to enter into the utility function for each choice and yields choice rules which are similar to those of regular

economic demand systems. Unfortunately, this model becomes intractable very quickly. A more modest modification is to incorporate alternative-specific constants into V_{ij} as a means of correcting for misspecification bias due to unobserved attributes. However, this does not help much in modeling substitutability *per se*.

An alternative, more parsimonious, approach is suggested by the family of substitution models having the general form:

$$P_j = \frac{R_j \exp(V_j)}{\sum_k R_k \exp(V_k)} \quad (11)$$

where,

R_j is a measure of the average degree of dissimilarity between the observed attributes of alternative j and the other alternatives.

The rationale for this approach is that the degree to which an alternative possesses distinctive properties affects its chances of being included in a restricted choice set, when the number of alternatives is large. Whether it is affected positively or negatively is an empirical question.

Several formulations have been suggested to measure an alternative's dissimilarity to other alternatives. Batsell (1981), for example, suggests the following:

$$R_j = \exp \left[\frac{1}{J-1} \sum_k \sum_m \phi_m |x_{jm} - x_{km}| \right] \quad (12)$$

where x_{jm} is the value of attribute m on alternative j ($m=1, \dots, M$) and ϕ_m is a substitution parameter reflecting the contribution of dissimilarity on the m th attribute to j 's overall dissimilarity. Alternatives that are very similar to others, in terms of their attributes that affect choice, will have low scores on R_j ; alternatives that are very unusual will have high scores.

Meyer and Eagle (1982) provide a slightly different measure where,

$$R_j = \left[\frac{1}{J-1} \sum_k 0.5 |r_{jk} - 1| \right]^\phi \quad (13)$$

and r_{jk} is the correlation coefficient between the attributes of alternatives j and k . The degree to which k differs from j is averaged across all alternatives k . Borgers and Timmermans (1987) suggest a similar formulation where,

$$R_j = \prod_m \left[\frac{1}{J-1} \sum_k |x_{jm} - x_{km}| \right]^{f_{m/M}} \quad (14)$$

where differences on one attribute are averaged across all alternatives and the geometric mean of these values over all attributes is calculated.

Fotheringham's competing destinations model (Fotheringham, 1983, 1986, 1988) is an explicitly spatial substitution model, where

$$R_j = \left[\frac{1}{J-1} \sum_k w_k / d_{jk} \right]^\phi \tag{15}$$

w_k is a weight (usually size). The term inside the brackets is a measure of the proximity of one spatial alternative to all the others and will be large for more central alternatives. This formulation has the advantage of being able to model hierarchical choice, where the hierarchy of choices is unknown (Fotheringham, 1988). This point is expanded below. The formulation for R_j in Equation (15), also allows more flexible spatial proximity or substitution variables to be defined. For instance, in store choice situations, some stores may be preferred to others because of their proximity to similar stores (comparative shopping), different stores (multi purpose shopping) or other activities such as banking facilities, restaurants, etc. (multi purpose trip-making). In each case, the formula for R_j can be adjusted to measure the proximity of a store to different land uses.

An alternative way of viewing the verifiable R_j in Equation (11) is a measure of the likelihood that alternative j is in a restricted choice set formed by an individual faced with too many alternatives to evaluate. Clearly, if every alternative is in the choice set evaluated by the individual, R_j is a constant and Equation (11) is equivalent to the MNL formulation.

Two assumptions made in the theoretical derivation of the competing destinations model separate it from other choice models derived and applied mainly in aspatial choice contexts. One is that there is a limit to an individual's ability to process large amounts of information and, therefore, spatial choice is likely to result from a hierarchical strategy, whereby, the alternatives are first partitioned. The second, is that due to the continuous nature of space, the composition of spatial clusters perceived by individuals is often uncertain and so a probability of cluster membership has to be attached to each alternative. Formally, consider that an individual evaluates only a subset L of the J alternatives in C' . Denote the probability of an alternative j being in the set L selected by individual i as $P_i(j \in L)$. Then in Equation (3) each alternative's utility should be weighted by the probability of that alternative being in the restricted choice set as follows:

$$P_i(j|C') = P_i [U_{ij} - U_{ik} + \ln P_i(k \in L)] > 0 \forall k \in C', j \neq k] \cdot P_i(j \in L) \tag{16}$$

which results in the more general choice model

$$P_i(j|C') = \frac{\exp(V_{ij}) \cdot P_i(j \in L)}{\sum_k \exp(V_{ik}) \cdot P_i(k \in L)} \tag{17}$$

and which indicates the role of R_j in the substitution models as a measure of the likelihood of j being in the restricted choice set formulated by individual i .

By incorporating a weight on each observable utility, the structure of the competing destinations model in Equation (17) is intrinsically different from that of the MNL model. The former no longer contains the undesirable (in spatial choice) IIA property: the ratio of the probabilities of selecting two alternatives is now:

$$\frac{P_i(j|C')}{P_i(k|C')} = \frac{\exp(V_{ij}) P(j \in L)}{\exp(V_{ik}) P_i(k \in L)} \tag{18}$$

which is no longer constant under the addition of new alternatives having differential effects on the values of $P_i(j \in L)$ and $P_i(k \in L)$.

The substitution models discussed above provide a tractable method of incorporating unequal substitutability when the choice set is large. However, it should be noted that such models do not account for taste variations across individuals. A general specification for modeling heterogeneity in the population within a logit framework is:

$$V_{ij} = X_{ij}\beta_1 + s_1 \alpha_j + (x_{ij}, S_i)\beta_2 \tag{19}$$

- where, β_1 is a $M \times 1$ vector of parameters measuring the effect of the M attribute variables;
- α_j is a $H \times 1$ vector of parameters measuring the effect of the H population characteristics in $s_j, j=1, \dots, J$;
- and β_2 is a $(M \times H) \times 1$ parameter vector for the interaction between attribute x_{ij} and characteristic s_{ih} .

The specification in (19) is general in that it allows population characteristics to either shift the systematic portion of utility upward, via the α_j 's, or to change the slope for an attribute, via the interaction parameters. Ben-Akiva and Lerman (1985) suggest that the suitability of the IIA assumption for a choice context, and thus the applicability of MNL, depends, in part, on the degree to which heterogeneities in the population are accounted for in the systematic component. As a demonstration of this point, they consider the following hypothetical situation. A heterogeneous population is divided into two equal-sized groups, which are internally homogeneous. Each group faces a choice between shopping in the downtown area (the CBD) or a suburban mall (SUB1). The probabilities for the first group are $P_{CBD} = 0.95, P_{SUB1} = 0.05$. For the second group we have, say, $P_{CBD} = 0.05$ and $P_{SUB1} = 0.95$. The population shares are, therefore, 50 percent for each center. A second mall, SUB2, is built in another suburb. This

second mall is identical to the first in observed attributes (including location relative to the population). The MNL predicts, for the first group, a new set of probabilities: $P_{CBD} = 0.9048$ and $P_{SUB1} = P_{SUB2} = 0.0476$. For the second population sub-group, the MNL predicts $P_{CBD} = 0.0256$ and $P_{SUB1} = P_{SUB2} = 0.4872$. The combined population shares are predicted by MNL to be 46.52 percent for the CBD and 26.74 percent for each mall and are not too far off the intuitively correct shares of 50 percent for the CBD and 25 percent for each mall. Thus, because heterogeneity in the population is accounted for, the violation of IIA is not too severe when it comes to aggregating the selection probabilities. This is because IIA applies to individual selection probabilities, not to the population as a whole. The practical implication of this is that the severity of the restrictions imposed by IIA can be reduced by capturing as much of the heterogeneity of the population as possible in specifying the systematic component of utility. However, capturing heterogeneity, while desirable even when IIA holds, does impose burdens - for each individual level characteristic which is built into the linear predictor, the number of coefficients to be estimated increases by $J-1$ if interaction effects are ignored and by $(J-1+M)$ per individual characteristic if one includes interaction effects. Further, capturing heterogeneity by including characteristics of the population in the regressor set means that some parameters are alternative-specific.

V. Random Component Substitutability

The Nested Logit Model

McFadden (1978) has developed a class of choice models, based on the generalized extreme value distribution (GEV), which can avoid the IIA assumption and yet are still consistent with utility maximizing behavior. The advantage of the GEV class of choice models is that the random components of utility, the e_{ij} 's in Equation (3), may be correlated across alternatives so that a model can be generated which allows for alternatives to have different levels of substitutability across alternatives. A particular case of the GEV family which is computationally tractable is the nested logit model (NMNL).

The NMNL assumes that the choice process can be hierarchically structured along the dimensions of the choice set. For example, suppose that the process being modeled is the housing choice problem described previously. It is hypothesized that consumers first choose a neighborhood, j , and then a housing unit, k , within that neighborhood. Given this choice strategy, the selection probability for the joint choice of a housing unit k in neighborhood j is

$$P_{ijk} = P_{ij} P_{ikj} \quad (20)$$

where P_{ijk} is the probability of choosing unit k in neighborhood j ;
 P_{ij} is the marginal probability of choosing j ;
 and P_{ikj} is the conditional probability of k , given that j is chosen.

The NMNL selection probabilities are:

$$P_{ijk} = \frac{\exp(V_{ij} + V_{ik} - \delta I_{ij})}{\sum_n \exp(V_{in} + (1-\delta)I_{in})} \quad (21.a)$$

$$P_{ij} = \frac{\exp(V_{ij} + (1-\delta)I_{ij})}{\sum_n \exp(V_{in} + (1-\delta)I_{in})} \quad j=1, \dots, J; \quad (21.b)$$

$$P_{ikj} = \frac{\exp(V_{ik})}{\exp(I_{ij})} \quad k=1, \dots, M_j; \quad (21.c)$$

where $I_{ij} = \ln \sum_{k=1}^{M_j} \exp(V_{ik} / (1-\delta))$,

n is an index of regions,

M_j represents the number of units in J

and δ is a parameter to be estimated where $0 \leq \delta < 1$.

The NMNL can thus be viewed as a system of models, in which lower level choices are nested in the upper-level choices through the parameter, δ , on the inclusive value, I_{ij} . With the restriction imposed on the range of $\delta = 0$, the NMNL reduces to a MNL in which housing unit and neighborhood are chosen simultaneously. For $\delta = 1$, the decision-maker is inferred to first choose a neighborhood and then, conditional on the choice of a neighborhood, choose a unit. Values of δ outside the unit interval are inconsistent with utility-maximizing behavior. A δ estimated as outside the unit interval, is evidence of model misspecification. However, δ also can be interpreted as approximating the correlation, or similarity, between the random or unobservable components of utility of each set of dwellings in each neighborhood. thus, if $\delta=0$, the dwelling units are inferred to be uncorrelated in their unobserved attributes. In that instance, the IIA assumption holds and MNL is appropriate for the choice set. As $\delta \rightarrow 1$, the dwelling units within each neighborhood are perceived by the consumer to be increasingly similar in their unobserved attributes. In this instance, IIA does not hold and the choice process must be modeled as a two-stage or hierarchical process. The key advantage of the NMNL is that it facilitates the modeling of alternatives with very different degrees of substitutability. However, the NMNL imposes a particular structure of unequal substitutability. Thus, it is assumed that, regardless of the value of δ , IIA is an ap-

appropriate assumption for each set of dwellings within each neighborhood j , as can be seen from the conditional probabilities of (21.c). IIA is not, however, assumed for the choice between neighborhoods.

The inclusive value term, for the neighborhood probabilities, can be viewed as an additional element of the neighborhood-level systematic component, which adjusts for unequal substitutability across neighborhoods due to their content of dwelling units. In that context, it should also be noted that the inclusive value can be viewed as a measure of the heterogeneity of the attributes of housing units within each neighborhood. Thus, viewing neighborhoods as grouped alternatives with the housing units as elemental alternatives, the NMNL framework makes explicit the sequential choice process which is assumed when alternatives.

It has been noted that the NMNL bears some resemblance to the elimination-by-aspects (EBA) model proposed by Tversky (1972) as a means of addressing the IIA problem. The EBA assumes that a decision-maker evaluates a set of alternatives on the basis of a weighting of the aspects, or attributes, of the alternatives. Alternatives are eliminated on the basis of whether or not their aspects meet the decision-maker's criteria. Unequal substitutability is not a problem since aspects are weighted, not alternatives. However, Tversky's model is hard to apply (Maddala, 1984) and yields results which are difficult to distinguish from NMNL. Thus, the NMNL encompasses a broader range of behaviors than the MNL. Further, the dimensioning of choice sets by qualitative aspects such as CBD/suburban can be handled by NMNL rather than dummy variables, though this raises obvious questions concerning the structuring of the choice process and also the determination of relevant aspects.

The NMNL model would thus appear to offer a solution to some of the problems created by differential degrees of substitutability and it has found wide use in aspatial contexts. However, use of the NMNL is likely to be more limited in spatial choice than in aspatial choice because of the difficulties often encountered in defining spatial clusters. Consider, for example, a consumer's mental map of grocery stores in a large city, where there are no obvious clusters of stores. The consumer is likely to have an impression about different shopping neighborhoods, but where the boundaries of such neighborhoods occur may not be at all obvious to the modelers. In order to operationalize the NMNL model, cluster membership for each individual would have to be 'guessed' by the modeler. Even if this were possible, the problem of different degrees of substitutability within clusters would remain. Consequently, the use of the NMNL in spatial choice is restricted to those situations where clusters can be

easily identified; an example, being the choice of store within shopping malls.

Finally, it should be noted that the estimation of the NMNL often presents computational difficulties. The full information maximum likelihood (FIML) approaches can be difficult to implement since the likelihood function involves a ratio of parameters. These are notoriously slow to converge. An alternative to estimating all parameters simultaneously is to proceed sequentially, i.e., estimate the parameters of the attributes of the alternatives at the lowest level of the hierarchy, compute the inclusive values, and then estimate the parameters of the attributes of the next order of alternatives, and so on. This method should yield consistent estimates but will be inefficient by comparison with the full information approach (Anas, 1982; Haynes and Phillips, 1982). Further, this sequential estimation approach does not yield the appropriate hypothesis tests for the coefficients and requires a rather complex correction process (Amemiya, 1978). Recent software packages (HLOGIT available from the Transport Research Center at Macquaire University) have significantly eased these concerns.

The Multinomial Probit Model

An alternative to the GEV approach in modeling discrete choice is based on the assumption that the random components of utility in Equation 3 are distributed normal rather than Type I Extreme Value. This assumption leads to the multinomial probit model (MNP). Hausman and Wise (1978) develops a very general model in which an individual's tastes for the observable attributes of the alternatives are randomly distributed. The utility of alternative j for individual i is written as:

$$U_{ij} = \sum_m (\beta_m + \beta_{jm}) x_{ijm} + \sum_h \alpha_{hj} + s_{ih} + \varepsilon_{ij} \quad (22)$$

Each parameter β_{im} measures individual specific departures from the average weighting of attribute m , given by β_m . ε_{ij} is an individual-specific, alternative-specific random variable. Thus, the systematic and random components can be distinguished as follows:

$$U_{ij} = \text{Fixed} + \text{random parts} \quad (23)$$

$$= \left\{ \sum_m \beta_m x_{ijm} + \sum_h \alpha_{hj} s_{ih} \right\} + \left\{ \sum_m \beta_{im} x_{ijm} + \varepsilon_{ij} \right\}$$

If the β_{im} 's and ε_{ij} are normally distributed then the random part of Equation 23 is also randomly distributed. In spatial choice, this modeling approach has the intuitively appealing feature that, if an individual has a higher than average valuation of one attribute, say travel time in

destination choice, then he/she will have a higher than average utility for all destinations with short travel times. This implies that there will be a covariance between the unobservable components of the utility of alternatives j and j' , which allows varying degrees of substitution between the alternatives.

However, this generality is not without cost. Because the cumulative probability distribution for the normal cannot be expressed in closed algebraic form, as it can for the generalized extreme value models, it must be computed numerically. This requires $J-1$ fold integration:

$$P_{ij} = \int_{-\infty}^{V_{i1} - V_{ij}} \dots \int_{-\infty}^{V_{i,j-1} - V_{ij}} \phi_{J-1}[\xi_1, \dots, \xi_{j-1} | \Omega_j] d\xi_1 \dots d\xi_{j-1} \quad (24)$$

where ϕ_{j-1} is the $J-1$ -variate normal density function with mean 0 and the variance covariance matrix, Ω_j , given by:

$$\Omega_j = \begin{bmatrix} \sum_m \alpha_m^2 [x_{im}^2 - x_{jm}^2] + \alpha_i^2 + \alpha_j^2 - 2\alpha_i \alpha_j, \dots \\ \dots \\ \sum_m \alpha_m^2 [x_{i,j-1,m}^2 - x_{j,m}^2] + \alpha_{i,j-1}^2 + \alpha_j^2 - 2\alpha_{i,j-1} \alpha_j \end{bmatrix} \quad (25)$$

Several different strategies have been taken in this numerical computation. The brute force approach is to integrate repeatedly Equation (24) by parts and develop a recursion formula for the integral. Even when only trivariate integration is required, this is usually computationally infeasible. As an alternative, Daganzo (1979) suggests using an approximation due to Clark (1961). Unfortunately, the error associated with this approximation is not well understood and is not easily controlled. CPU times for this approach go up in proportion to the square of the size of the choice set. While this sounds large, it is quite attractive compared to other techniques which tend to go up geometrically with the size of the choice set. Applications with this technique have been quite mixed, particularly when some of correlations are negative. This suggests that it be used predominantly for exploratory work.

Lerman and Manski (1982) use Monte Carlo techniques to estimate the integral. A large set of $J-1$ variate normal deviates are generated. The fraction of these observations for the chosen alternative is an approximation for the probability in Equation (24). This approach works well as long as none of the probabilities are very rare. It does have the drawback that the likelihood function is merely estimated and the statistical properties of this sampling error are typically ignored. Sickles and Taubman (1986), develop Gaussian techniques for approximating 10-fold integrals. Unlike the Clark algorithm, precision is improved by increasing the number of points at which the density function is evaluated.

The MNP, though a more general model than MNL or NMNL, has only rarely been applied in a spatial choice

context. A primary reason for this lies in the computational difficulties associated with the integration for Equation (24). However, in applications, restrictions are typically imposed on Ω_j for identification purposes and to reduce the computational burden. For example, restricting Ω_j to $\sigma^2 I$, where I is the identity matrix, leads to independent probit and, in effect, assumes IIA. Miller and Lerman (1979, 1981) have developed a probit model for the joint choice of a retail location and store size. Their model imposes a set of restrictions on Ω_j which are similar to the restrictions implied by a NMNL under a choice strategy by which location is determined prior to the selection of a store size category. Miller and Lerman's model was estimated for 18 alternatives using Chomp, an algorithm which employs Clark's approximation. Though some difficulty was experienced in estimating parameters, Miller and Lerman (1981) suggest that MNP is feasible even with a large number of discrete alternatives. However, this claim is predicated on the use of an approximation which is suspect. Kamakura and Srivastava (1982) propose a model with a full variance-covariance matrix which, however, only requires two additional parameters. This model is similar to the MNL substitution models described by Equations (11) in that it assumes that unequal substitutability between alternatives can be modeled as a function of distances between the observed attributes, though Kamakura and Srivastava use Euclidean distance. An application of their model to spatial choice analysis, again using Clark's approximation, can be found in Borgers and Timmermans (1987).

Obviously, if a model robust to IIA is estimated such as MNP or one of the generalized extreme value (GEV) models, Wald tests are appropriate. Several different types of specification tests for IIA are available and are reviewed in Haynes, Good and Dignan (1988).

VII. Concluding Remarks

As with most models developed in an aspatial environment, the introduction of space into discrete choice models is problematic. Spatial choice is not merely "aspatial choice with the addition of a distance variable": space adds a complexity to the relationships between alternatives that is not found in aspatial choice. In spatial choice, alternatives have fixed and unique locations vis-a-vis each other, which affects the degree to which alternatives are perceived as substitutes for each other.

This paper has argued that for a variety of reasons spatial choice is particularly susceptible to violation of IIA. However, the solutions to this problem generally employed in aspatial choice (use of the nested logit model, for example), are not as useful in spatial choice. It is argued that solutions lie in the generation of an explicitly spatial

theory of choice from which particular model forms can be generated. In this manner, the competing destinations model in which the degree of substitution between alternatives is continuous, rather than discrete, is seen as a step in the right direction.

Notes

¹The Type I Extreme Value distribution (Fisher and Tippet, 1928) is the double negative exponential distribution which is sometimes incorrectly referred to as the Weibull distribution.

References

- Amemiya, T., "On a two-step estimation of a multivariate logit model", *Journal of Econometrics*, 8, 13-21, 1978.
- Anas, A., *Residential Location Markets and Urban Transportation: Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models*, New York: Academic Press, 1982.
- Batsell, R. R., "A Multiattribute Extension of the Luce Model Which Simultaneously Scales Utility and Substitutability, Working paper from a conference at Quail Roost, 1981.
- Ben-Akiva, M., and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, Mass.: MIT Press, 1985.
- Borgers, A. and H. Timmermans, "Choice model specification, substitution, and spatial structure effects: a simulation experiment", *Regional Science and Urban Economics*, 17, 29-47, 1987.
- Brown, H. P., "A Model of Weekend Recreational Travel Demand, Working Paper RP/06/79/Trans, Department of Civil Engineering, University of Melbourne, Australia, 1979.
- Carlton, D. W., "The location and employment choices of new firms: an econometric model with discrete and continuous endogenous variables", *The Review of Economics and Statistics*, 65, 440-449, 1983.
- Clark, E., "The greatest of a finite set of random variables", *Operations Research*, 9, 145-162, 1961.
- Clark, W. A. V., and J. L. Onaka, "An empirical test of a joint model of residential mobility and housing choice", *Environment and Planning A*, 17, 915-930, 1985.
- Daganzo, C. F., *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, New York: Academic Press, 1979.
- Dignan, T., K. E. Haynes, D. Conway and M. Shrestha, "Land and Landlessness Among Rural-to-Rural Migrants in Nepal's Tarai Region", *International Regional Science Review*, 12, forthcoming, 1988.
- Domenciah, T., and D. McFadden, *Urban Travel Demand: A Behavioral Analysis*, Amsterdam: North-Holland, 1975.
- Fischer, M. M., and G. Maier, "Regional labour supply mobility and the logit model, testing some alternative probability choice structures by means of Monte Carlo simulation." Paper Symposium IGU Working Group on Systems and Mathematical Models' Besamcon, 1984.
- _____, and P. Nijkamp, "Developments in explanatory discrete spatial data and choice analysis", *Progress in Human Geography*, 515-551, 1985.
- Fisher, R.A. and L.H.C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample", *Proceedings of the Cambridge Philosophical Society* 24, 180-190, 1928.
- Fotheringham, A. S., "A new set of spatial-interaction models: the theory of competing destinations", *Environment and Planning A*, 15, 15-36, 1983.
- _____. "Modelling hierarchical destination choice", *Environment and Planning A*, 18, 401-418, 1986.
- _____. "Consumer store choice and choice set definition", *Marketing Science*, 7, 299-310, 1988.
- _____, and M. E. O'Kelly, "Spatial Interaction Models: Formulations and Applications", Kluwer Academic: Boston, 1989.
- Gale, S., "Inexactness, fuzzy sets, and the foundations of behavioural geography", *Geographical Analysis*, 4, 337-349, 1972.
- Gaudry, M. J. I., and M. G. Dagnais, "The dogit model", *Transportation Research*, 13B (2), 105-112, 1979.
- Gould, P. and R. White, *Mental Maps*, Boston: Allen and Unwin, 1974.
- Greene, W. H., LIMDEP, Dept. of Economics, Graduate School of Business Administration, New York University, 1985.
- Hansen, E. R., "Industrial location choice in Sao Paulo, Brazil: a nested logit model", *Regional Science and Urban Economics*, 17, 89-108, 1987.
- Hausman, J., and D. McFadden, "Specification tests for the multinomial logit model", *Econometrica*, 52 (5), 1219-1240, 1984.
- _____, and D. Wise, "A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences", *Econometrica*, 46 (2), 403-426, 1978.
- Haynes, K. E., and F. Y. Phillips, "Constrained Minimum Discrimination Information: A Unifying Tool for Modeling Spatial and Individual Choice Behavior", *Environmental and Planning A*, 14, 1341-1354, 1982.
- _____, D. H. Good and J. Dignan, "Discrete Spatial Choice and the Axiom of Independence from Irrelevant Alternatives", *Socio-Economic Planning Sciences*, 22(b), 241-251, 1988.
- Hensher, D. A., and L. W. Johnson, *Applied Discrete Choice Modelling*, New York: Halstead Press, 1981.
- Hervitz, H. M., "An Inquiry into the Determinants of Inter-Regional Migration within the Framework of a Multinomial Logit model, Unpublished Ph. D. Dissertation, Indiana Uni-

- versity, Dept. of Economics, 1983.
- Horowitz, J., "Specification tests for nested logit models", *Environment and Planning A*, 19, 395-402, 1987.
- Huff, D. L., Determination of Intraurban Retail Trade Areas, Los Angeles: Real Estate Research Program, University of California, Los Angeles, 1962.
- Kamakura, W. A., and R. K. Srivastava, "Predicting brand shares under conditions of brand interdependence", *Journal of Marketing Research*, 21, 420-434, 1984.
- Keifer, N., "Testing for dependence in multivariate probit models", *Biometrika*, 69, 161-166, 1982.
- Lerman, S. and C. Manski, "On the use of simulated frequencies to approximate choice probabilities", in C. Mansky and D. McFadden (eds.), *Structural Analysis of Discrete Data: With Econometric Applications*, Cambridge, MA.: M.I.T. Press, 1982.
- Linneman, P., and P. E. Graves, "Migration and job change: a multinomial logit approach", *Journal of Urban Economics*, 14, 263-279, 1983.
- Luce, R. D., Individual Choice Behavior, New York: J. Wiley & Sons, 1959.
- Maddala, G. S., Limited-Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press, 1983.
- Manski, D. G., "The structure of random utility models", *Theory and Decision*, 8, 229-254, 1977.
- McFadden, D., "Conditional logit analysis of qualitative choice behavior", in P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press, 105-42, 1974.
- _____, "Modeling the choice of residential location", in A. Karlquist, L. Lundquist, F. Snickers, and J. L. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, Amsterdam: North Holland, 75-96, 1978.
- _____, "Econometric models of probabilistic choice", in C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data With Econometric Applications*, Cambridge, MA: MIT Press, 198-272, 1981.
- _____, "Econometric analysis of qualitative response models", in Z. Griliches and M. D. Intriligator, *Handbook of Econometrics*, Vol. II, Amsterdam: North-Holland, 1395-1457, 1984.
- Meyer, R. J., and T. C. Eagle, "Context-induced parameter instability in a disaggregate-stochastic model of store choice", *Journal of Marketing Research*, 19, 62-71, 1982.
- Miller, E. J., and S. R. Lerman, "A model of retail location, scale, and intensity", *Environment and Planning A*, 11, 177-192, 1979.
- _____, and S. R. Lerman, "Disaggregate modelling and decisions of retail firms: a case study of clothing retailers", *Environment and Planning A*, 13, 729-746, 1981.
- Mueller, C. F., The Economics of Labor Migration: A Behavioral Analysis, New York: Academic Press, 1982.
- Odland, J., and M. Ellis, "Disaggregate migration behavior and the volume of interregional migration", forthcoming in *Economic Geography*, 1987.
- Onaka, J., and W. A. V. Clark, "A disaggregate model of residential mobility and housing choice", *Geographical Analysis*, 15 (4), 287-304, 1983.
- Pipkin, J. S., "Fuzzy sets and spatial choice", *Annals of the Association of American Geographers*, 68, 196-204, 1978.
- Quigley, J. M., "Housing demand in the short run: an analysis of polytomous choice", *Exploration in Economic Research*, 3, 76-102, 1976.
- _____, "Consumer choice of dwelling, neighborhood, and public services", *Regional Science and Urban Economics*, 15, 41-63, 1985.
- Shepherd, E. S., "Theoretical underpinnings of the gravity hypothesis", *Geographical Analysis*, 10 (4), 386-401, 1978.
- Sickles, R. C., and P. Taubman, "An analysis of the health and retirement status of the elderly", *Econometrica*, 54, 1339-1356, 1986.
- Swait, J., and M. Ben-Akiva, "Incorporating random constraints in discrete choice models: and application to mode choice in Sao Paulo, Brazil", *Transportation research B*, 1984.
- Transport Research Center, "HLOGIT" Software Program, Macquarie University, 1989.
- Tversky, A., "Elimination by aspects: a theory of choice", *Psychological Review*, 79, 281-299, 1972.
- Williams, H. C. W. L., "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", *Environmental and Planning A*, 9, 285-344, 1977.
- Wilson, A. G., J. D. Coelho, S. M. Macgill and H. C. W. L. Williams, "Optimization in Locational and Transport Analysis Chichester, U. K.: John Wiley, 1981.
- Wrigley, N. "Quantitative Methods: Developments in Discrete Choice Modeling," *Progress in Human Geography*, 6, 547-562, 1982
- _____, *Categorical Data Analysis for Geography and Environmental Scientists*, UK: Longman 1985
- _____, and P.A. Longley "Discrete Choice Modeling in Urban Analysis," 45-94 in Herbert and R.J. Johnston (eds.) *Geography and the Urban Environment*, 6, U.K.: Wiley 1984.
- Zadeh, L. A., "Fuzzy sets", *Information and Control*, 8, 338-353, 1965.